

ATTN: Ed Hart. 66252

Access DB# _____

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Celine Qian Examiner #: 75710 Date: 5/8/02
Art Unit: 1636 Phone Number 306-0283 Serial Number: 09/662128
Mail Box and Bldg/Room Location: 11E 051 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Modified cre recombinase gene for mammals
Inventors (please provide full names): Shuji Hiyagawa, Masaru Okabe

Earliest Priority Filing Date: 9/17/1999

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

search SEQ ID NO: 1 (both commercial & interference)

search nucleic acid sequence encoding amino acid
sequence of SEQ ID NO: 2

RECEIVED
MAY - 8 2002
(STIC)

STAFF USE ONLY

Type of Search		Vendors and cost where applicable
Searcher: _____	NA Sequence (#) <u>1</u>	STN _____
Searcher Phone #: _____	AA Sequence (#) <u>1</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>5/8/02</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>5/14/02</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>05</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: _____	Other _____	Other (specify) _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2002, 00:54:43 ; Search time 253.13 Seconds
(without alignments)
7121.875 Million cell updates/sec

Title: US-09-662-128a-1
Perfect score: 1050
Sequence: 1 atgcccaagaagaagagaa.....gctctgagagacgagc 1050

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1050	100.0	22	AAH41175	Mammalian Cre reco
2	755	71.9	22	AAH41175	Chimeric Cre reco
3	754.4	71.8	22	AAH41175	Chimeric Cre reco
4	639.4	60.9	22	AAH41175	Chimeric Cre reco
5	639.4	60.9	22	AAH41175	Chimeric Cre reco
6	639.4	60.9	22	AAH41175	Chimeric Cre reco
7	636.4	60.6	22	AAH41175	Chimeric Cre reco
8	630	60.0	22	AAH41175	Chimeric Cre reco
9	621.8	59.2	21	AAH41175	Chimeric Cre reco

10	617.4	58.8	1074	22	AAH41175	TAT-Cre fusion DNA
11	617.4	58.8	1623	22	AAH41175	Delta VP22Cre-Stre
12	617.4	58.8	1740	21	AAH41175	GST-Cre coding reg
13	617.4	58.8	2004	18	AAH41175	Chimeric Cre-LDB-G
14	617.4	58.8	2004	18	AAH41175	VP22-Cre fusion DN
15	617.4	58.8	2055	22	AAH41175	VP22CreStreptag fu
16	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
17	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
18	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
19	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
20	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
21	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
22	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
23	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
24	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
25	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
26	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
27	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
28	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
29	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
30	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
31	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
32	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
33	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
34	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
35	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
36	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
37	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
38	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
39	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
40	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
41	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
42	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
43	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
44	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G
45	617.4	58.8	2346	22	AAH41175	Chimeric Cre-LDB-G

ALIGNMENTS

RESULT 1	
AAH41175	AAH41175 standard; DNA; 1050 BP.
AC	AAH41175;
AC	23-AUG-2001 (first entry)
DT	Mammalian Cre recombinase gene.
XX	Mammalian Cre recombinase; organ transplantation; gene therapy;
XX	animal disease model; ds.
XX	Mammalia.
OS	JP2001086989-A.
PN	03-APR-2001.
XX	17-SEP-1999; 99JP-0264364.
XX	17-SEP-1999; 99JP-0264364.
XX	17-SEP-1999; 99JP-0264364.
XX	(OSAU) UNIV OSAKA.
XX	WPI; 2001-360321/38.
DR	P-PSDB; AAB98695.
XX	Mammalian type Cre recombinase gene modified so as to have a high
PT	expression efficiency in a mammal by selecting a codon of high
PT	frequency for use in the mammal for expressing Cre recombinase -
XX	

PS Claim 2; Page 7-9; 12pp; Japanese.

XX
CC The present sequence is a mammalian type Cre recombinase gene. This
CC sequence can be modified so as to have a high expression efficiency in a
CC mammal by selecting a codon of high frequency for use in the mammal for
CC expressing Cre recombinase. The gene can be used in organ
CC transplantation, gene therapy and creation of animal disease models.
XX

SQ Sequence 1050 BP; 204 A; 383 C; 339 G; 124 T; 0 other;

Query Match 100.0%; Score 1050; DB 22; Length 1050;
Best Local Similarity 100.0%; Pred. No. 2,3e-145;
Matches 1050; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atgcccaagaagaagaagtagcaacactgtgacccgtgacacgaacactgcccgc 60
DB 1 atgcccaagaagaagaagtagcaacactgtgacccgtgacacgaacactgcccgc 60
QY 61 ctgcccgtgagacgacacacgacgagtgctgcaagaacctgatgatgttccgcgac 120
DB 61 ctgcccgtgagacgacacacgacgagtgctgcaagaacctgatgatgttccgcgac 120
QY 121 cgcgaagccttcacgagacacacttgtagaatgtgtgtgctgagcgtgctgagcgc 180
DB 121 cgcgaagccttcacgagacacacttgtagaatgtgtgtgctgagcgtgctgagcgc 180
QY 181 gctctgtgcaagctgcaacacacgcaatgtgttcccgacgacccgagagacgtgcgac 240
DB 181 gctctgtgcaagctgcaacacacgcaatgtgttcccgacgacccgagagacgtgcgac 240
QY 241 taactgtctgtactcgcagagccgcgctgagcgtgcaagaacctcagcagacactgggc 300
DB 241 taactgtctgtactcgcagagccgcgctgagcgtgcaagaacctcagcagacactgggc 300
QY 301 cagctggaacatgctgcacacgacgagcgtgctgcccgcacgacgacgaacgacgctg 360
DB 301 cagctggaacatgctgcacacgacgagcgtgctgcccgcacgacgacgaacgacgctg 360
QY 361 agcctgtgtagtcgcgcacatccgcgaagaagaaacgtgacgcgcgacgcgcgaacgac 420
DB 361 agcctgtgtagtcgcgcacatccgcgaagaagaaacgtgacgcgcgacgcgcgaacgac 420
QY 421 gctctgtgtagtcgcgcacatccgcgaagaagaaacgtgacgcgcgacgcgcgaacgac 480
DB 421 gctctgtgtagtcgcgcacatccgcgaagaagaaacgtgacgcgcgacgcgcgaacgac 480
QY 481 cgtctgcagagacatccgcgaacactgtgcttctggtgacatccacacacccctgtgcgc 540
DB 481 cgtctgcagagacatccgcgaacactgtgcttctggtgacatccacacacccctgtgcgc 540
QY 541 atcgccgagatcgccgcacatccgcgtgaaagacatcagccgacacgacgacgacgacgac 600
DB 541 atcgccgagatcgccgcacatccgcgtgaaagacatcagccgacacgacgacgacgacgac 600
QY 601 ctgatacaacatccgcgcacacacacactgtgtagacacgacgacgacgacgacgacgac 660
DB 601 ctgatacaacatccgcgcacacacacactgtgtagacacgacgacgacgacgacgacgac 660
QY 661 agcctgtgtagtcgcgcacatccgcgtgaaagacatcagccgacgacgacgacgacgac 720
DB 661 agcctgtgtagtcgcgcacatccgcgtgaaagacatcagccgacgacgacgacgacgac 720
QY 721 ccccaacaactactgttctgtcgcgcgtgacgaagaacgacgacgacgacgacgacgac 780
DB 721 ccccaacaactactgttctgtcgcgcgtgacgaagaacgacgacgacgacgacgacgac 780
QY 781 agcgaagttagcagccgacgacgacgacgacgacgacgacgacgacgacgacgacgac 840
DB 781 agcgaagttagcagccgacgacgacgacgacgacgacgacgacgacgacgacgacgac 840
QY 841 ggcgcgaagaagacagcagcgcgacgacgacgacgacgacgacgacgacgacgacgacgac 900
DB 841 ggcgcgaagaagacagcagcgcgacgacgacgacgacgacgacgacgacgacgacgacgac 900

```

```

DB 841 ggcgcgaagaagacagcagcgcgactactggtgagcgcacacgacgacgacgacgac 900
QY 901 ggcgcgcgcgcgcacacatgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
DB 901 ggcgcgcgcgcgcacacatgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
QY 961 gctctgacacacgacgacacatcgtgatgaactacatccgaacctgagacgacgacgac 1020
DB 961 gctctgacacacgacgacacatcgtgatgaactacatccgaacctgagacgacgacgac 1020
QY 1021 gccatgtgctgctgctgctgctgagagacgacgacgacgacgacgacgacgacgac 1050
DB 1021 gccatgtgctgctgctgctgagagacgacgacgacgacgacgacgacgacgacgac 1050

RESULT 2
AADI0217
ID AADI0217 standard; DNA; 2346 BP.
AC AADI0217;
XX
AC AADI0217;
XX
DE 24-SEP-2001 (first entry)
XX
DE Chimeric recombinase DNA encoding moCre:FLPm protein.
XX
KW Site specific recombinase; expression cassette; chimeric;
KW moCre:FLPm protein; ds.
OS Chimeric - Saccharomyces sp.
OS Chimeric - Bacteriophage P1.
OS Chimeric - Zea mays.
XX
FH Key
FT Location/Qualifiers
FT CDS 1..2346
FT /tag= a
FT /product= "Chimeric moCre:FLPm protein"
PN US6262341-B1.
PD 17-JUL-2001.
XX
PD 17-JUL-2001.
XX
PF 17-NOV-1998; 98US-0193503.
XX
PR 18-NOV-1997; 97US-0065613.
PR 18-NOV-1997; 97US-0065627.
PR 08-SEP-1998; 98US-0099435.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Baszczynski CL, Lyznik LA, Gordon-Kamm WJ, Guan X, Rao AG;
PI Tagliani LA;
DR WPI; 2001-450495/48.
DR P-PSDB: AAE05412.
XX
PT Integrating DNA of interest into genome of eukaryotic cell, by
PT transforming plant cell with transfer cassette comprising DNA flanked
PT by target sites for site-specific recombinases and providing
PT recombinases in cell
XX
PS Claim 4; Column 23-28; 30pp; English.
XX
CC The invention relates to compositions and methods for introducing
CC a DNA of interest into a genomic target site. The methods and
CC compositions involve the use of a combination of target sites for two
CC site specific recombinases and expression of a chimeric recombinase
CC with dual target site specificity. The compositions comprise novel
CC site-specific recombinases with specificities to multiple target sites,
CC and nucleotide sequences and expression cassettes encoding these
CC recombinases or target sites. The method of integrating foreign DNA
CC into genome of eukaryotic cell involves transforming the cell having
CC target sites for the novel recombinase with a DNA of interest that is
CC flanked by corresponding target sites. The method is useful for

```


[illegible]

RESULT	5
AAAD09269	
ID	AAAD09269 standard; DNA; 4960 BP.
XX	
AC	AAAD09269;
XX	
DT	12-SEP-2001 (first entry)
XX	
DE	PCMV-I-Cre-PA vector DNA.
XX	
XX	
KM	DNA recombinase domain; protein transduction domain; PMD;
KM	gene alteration; fusion protein; Human immunodeficiency virus;
KM	HIV; PCMV-I-Cre-PA vector; ds.

XX Unidentified.
OS
XX
XX WO200149832-A2.
PN
XX
XX
PD 12-JUL-2001.
PF
XX 05-JAN-2001; 2001WO-EP00060.
XX
XX 07-JAN-2000; 2000EP-0100351.
PR 10-NOV-2000; 2000EP-0124595.
XX
XX
PA (ARTE-) ARTEMIS PHARM GMBH.
XX
XX
PI Schwenk F;
XX
XX WPI; 2001-441873/47.
DR
XX
XX
XX
XX Using site-specific DNA recombinase domain/protein transduction domain
PI fusion proteins for inducing target gene alterations in organisms or
XX cell cultures -
XX
XX Example; Page 62-63; 85pp; English.
XX

CC The present invention relates to use of fusion proteins comprising
CC a site-specific DNA recombinase domain e.g. Cre and a protein
CC transduction domain (PTD) e.g. the Human immunodeficiency virus
CC (HIV) derived TAT peptide, for preparing an agent for inducing
CC target gene alterations in a living organism or cell culture. The
CC present invention also provides a method for inducing gene
CC alterations in living organisms using the fusion proteins of the
CC invention. The present sequence is a pCMV-T-Cre-pA vector DNA.
XX
SQ Sequence 4360 BP; 1225 A; 1213 C; 1296 G; 1226 T; 0 other;

Query Match	60.9%;	Score 639.4;	DB 22;	Length 4960;
Best Local Similarity	75.6%;	Pred. No. 1.8e-85;		
Matches 793; Conservative	0;	Mismatches 256;	Indels 0;	Gaps 0.

[illegible]

```

Db      1457  cgcgcgcagatatacgtatctgcgcatctctgggattctgtatataccacctgtacgt 1516
Oy      541  atcgccagatcgcccgccgctggaagacatcagccgacgagcgccgcatg 600
Db      1517  atagccgaatctgcagagacaggttaagatatctacgtactgacggttgaggagt 1576
Oy      601  ctgactacatcgccgacacgaagccctggtagcaccgcccggctggagaagccctg 660
Db      1577  ttaatccattctggcagaaacgctgttagcacccgcaagtgtagaagaagcactt 1636
Oy      661  agcctggcgctgacacgctggtggagcgttgatcaagcgttgagcgctggccgacgac 720
Db      1637  agcctgggggtaactaaacggtctcgagcagatcttcgctcctggtgtagctgagt 1696
Oy      721  cccaacaactactgtcttcgcgcgtgcaagaacgcgctggccgcccacgaccac 780
Db      1697  ccgaataactactcgttttgcgcggtcagaaaaatgtgtgtgcgcgcacatctgccc 1756
Oy      781  agccagctgagacccggcccttggaagcgtatcttcgaagccacccacgctgactac 840
Db      1757  agccagctatcaactcgcgccttggaaggtattcttgaaacacacccatctgattac 1816
Oy      841  ggcgcgaagacgacgagcgccacgctacactgctgagcgccacgacgcccgtg 900
Db      1817  ggcgctaaagatgactcgtgcagagatacctgctgctgctgacacagtgcccgtgc 1876
Oy      901  ggcgcgcgcgcgacatgcccgcgcgcgctgagcatcccccagatcagcagcgccgc 960
Db      1877  ggaagcgcgagatgaltgcccgcgtctgagttcaataccgagacatcagcagcggt 1936
Oy      961  ggcctggacaaagtgtaacatcgtagactacatccgcaacctgacagcagcgagc 1020
Db      1937  ggcctggacaaatgtataatgtctacatatactcgttaaccctgagtagtgaacagg 1996
Oy      1021  gccatggtgcgcctgctgagagcgca 1049
Db      1997  gcaatggtgcgcctgctgagagtgca 2025

RESULT 6
AAD04928 standard; DNA; 5365 BP.
AC      AAD04928;
DT      17-JUL-2001 (first entry)
DE      Retroviral vector PBABE-pgkCre used as gene trap in embryonic stem cells.
XX      Gene trapping construct; conditional mutation; unidirectional inversion;
KW      recombinase recognition sequence; RRS; disruption cassette;
KW      selection cassette; transgenic organism; retroviral vector;
KW      PBABE-pgkCre; cre recombinase; phosphoglycerate kinase; PGK promoter; ds.
XX      Chimeric - Moloney murine leukemia virus.
OS      Chimeric - Mus sp.
OS      Chimeric - Bacteriophage P1.
OS      Chimeric - Rhesus macaque polyoma virus.
XX      Key Location/Qualifiers
FH      LTR 8..480
FT      /*tag= a
FT      /note= "5' partial MMLV U3, MMLV R and MMLV U5"
FT      /partial
FT      8..335
FT      /*tag= b
FT      /note= "5' partial MMLV U3"
FT      repeat_unit 336..402
FT      /*tag= c
FT      /rpt_type= DIRECT
FT      /note= "5' MMLV R direct repeat"
XX

```

```

FT      misc_feature 403..480
FT      /*tag= d
FT      /note= "5' MMLV U5"
FT      misc_feature 481..1374
FT      /*tag= e
FT      /note= "MMLV primer binding site and extended
FT      packaging signal"
FT      1417..1921
FT      /*tag= f
FT      /note= "Mouse phosphoglycerate kinase (PGK) promoter"
FT      1972..3024
FT      CDS
FT      /*tag= g
FT      /product= "Bacteriophage P1 Cre recombinase protein with
FT      a simian virus 40 large T-antigen N-terminal nuclear
FT      localisation signal"
FT      3088..3168
FT      /*tag= h
FT      /note= "Promoter/enhancer deleted MMLV U3"
FT      3187..3332
FT      LTR
FT      /*tag= i
FT      /note= "3' MMLV R and MMLV U5"
FT      repeat_unit 3187..3253
FT      /*tag= j
FT      /rpt_type= DIRECT
FT      /note= "3' MMLV R direct repeat"
FT      3254..3332
FT      misc_feature
FT      /*tag= k
FT      /note= "3' MMLV U5"
XX      WO200129208-A1.
XX      26-APR-2001.
XX      16-OCT-2000; 2000WO-EP10162.
XX      16-OCT-1999; 99EP-0120592.
XX      27-OCT-1999; 99US-0162016.
XX      (ARTE-) ARTEMIS PHARM GMBH.
XX      (FRAN-) FRANKGEN BIOTECHNOLOGIE AG.
XX      Kuehn R, Von Melchner H, Altschmied J;
XX      WPI; 2001-308486/32.
XX      DR      New gene trapping construct capable of causing conditional mutations in
XX      PT      genes, comprises functional DNA segment inserted in sense or antisense
XX      PT      direction relative to gene to be trapped
XX      PS      Example 1; Page 55-57; 78pp; English.
XX      The present invention relates to a conditional gene trapping construct
XX      capable of causing conditional mutations in genes. The gene trapping
XX      construct comprises two functional DNA segments, each being flanked by
XX      two recombinase recognition sequences (RRSs) specific to site specific
XX      recombinase which is capable of unidirectional inversion of double
XX      standard DNA segment. One of the DNA segment (disruption cassette) is
XX      inserted in antisense orientation relative to the transcriptional
XX      orientation of the gene to be trapped. The other DNA segment (selection
XX      cassette) is inserted in sense direction relative to the transcriptional
XX      orientation of the gene to be trapped. The cell comprising the gene
XX      trapping construct is useful for the identification and/or isolation of
XX      genes. The transgenic organism comprising the gene trapping construct is
XX      useful to study gene function at various developmental stages. The gene
XX      trapping construct is useful for mutationally inactivating all cellular
XX      genes. The present sequence is retroviral vector PBABE-pgkCre which is
XX      used as a gene trap in embryonic stem (ES) cells. The vector
XX      PBABE-pgkCre is constructed by inserting Bacteriophage P1 cre recombinase
XX      coding region with a simian virus 40 large T-antigen N-terminal nuclear
XX      localisation signal and mouse phosphoglycerate kinase (PGK) promoter into
XX      Moloney murine leukemia virus based vector PBABEpuro.

```

Sequence 5365 BP; 1208 A; 1448 C; 1402 G; 1307 T; 0 other;

Query Match 60.9%; Score 639.4; DB 22; Length 5365;
Best Local Similarity 75.6%; Pred. No. 1.8e-85;
Matches 793; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```

QY 1 atgcccaagaagaagatgagcaaccctgctgacccgagaccagaacctgcccgc 60
DB 1972 atgcccaagaagaagaagatgctcccaattacccagctacacaaatttgcctgca 2031
QY 61 ctgcccgtgagccgacccagcagaggtgacgaagaacctgagatgttccgcgac 120
DB 2032 ttaccggtcagatgacagatgagtgaggttcgcaagaacctgagatgttccgcgac 2091
QY 121 gcccagggcttcagagcaacacacctggaagatgtctgtgacggtgtccgcagctgggac 180
DB 2092 gcccagggcttcctcgtgacataccttggaataatgtctgtcgtctgtccggtcgtgagcg 2151
QY 181 gcttggtgcaagctgaacaacacgcaagtggtcccccgcgagcccgagagagctgcgac 240
DB 2152 gcatgtgcaagtgatgaataacgcggaatggttcccgagaaacctggaagtgtccgcgac 2211
QY 241 taacctggtacccctgacgagcccgccgctgagcgtgagacacatccagcaacctgggc 300
DB 2212 tatctctatctcagcgccgctgctgagtaaaatactacacagcaacatttggc 2271
QY 301 cagctgacacatgctgacccgcgagccgagccctgcccgcacacgacgaagacgcccgtg 360
DB 2272 cagctgacacatgctcactcactgctcgcgtccgagccacacgaagtgacgaatgctgtt 2331
QY 361 agctgtgtgacgcccgcacatccgcaagaagacgttgacgcccgcgagccgacgaacag 420
DB 2332 tcaactgggtatgctgagctgacccgaagaagacgttgacgcccgtgacgtgcaaacag 2391
QY 421 gacctggccttcgagccgacccgacttgacacaggtgacgacctgagaggaacagac 480
DB 2392 gctctagcgttcgacgacactgacttgcacaggttgccttcaactcaggaataatgagat 2451
QY 481 cgctgacgagacatccgcaacactgaccttccctgggacatcgccctacaacacctgtcgcc 540
DB 2452 cgctgacgagatatacgttaacttgcgacttccctgggaggttctataacacccgttaagt 2511
QY 541 atccgcaagatcgccgacccgctgctgaaagacatccgacccgacgagccgacgt 600
DB 2512 atagcgaataatgcaagatcagaaggttaaaagatattcactcagatgaggaagatg 2571
QY 601 ctgattcacatcgccgacacacacccctgtgagcagccgctgaggaagagccctg 660
DB 2572 ttaatcataattgcaagaacgaaacgctgtgacacccgaggtttagaggaagacatt 2631
QY 661 agcctggcgctgacgaagcgtggtgagcgtgacgacgagcggtgacgagcagac 720
DB 2632 agcctgggggttaactaaacatcgctgagcagatgattccgtctcctgtgagtgat 2691
QY 721 cccaacacatcctgttctcgcgctgacgaagaagcgtgagcccccacgacgac 780
DB 2692 ccgaataacactgttcttgcgcggtcagaaaaaaggtgtcgcgcgacatccatccacc 2751
QY 781 agccagctgagaccccgccgctgagggcatcttcgaagcccccacgacctgactac 840
DB 2752 agccagctacatcactcgccctgaggaagatttttgaagcaactcatcatgattac 2811
QY 841 ggcgcgaagagacagcgccgacgctactcctgctgagcgagcgacgagccgctg 900
DB 2812 ggcgcgaagagatgacctgtctgacgagatacctgctgctgacacagctgccgtgctc 2871
QY 901 ggcgcgcccgcgacatggccgagccgagcgtttagacatcccccagatcatgagccgac 960
DB 2872 ggcgcgcccgcgagatagtcgcccgcgctgaggttcaataccggagatcatgcaagctg 2931
QY 961 ggcctgacacagctgaacatcgtgatacatcatccgaacctgagacgagacgagc 1020
DB 1020 ggcctgacacagctgaacatcgtgatacatcatccgaacctgagacgagacgagc 1020

```

```

DB 2932 ggcgcgaacatgtaaatatgtcatgaactatccgtaacctgagtagtgaacaggg 2991
QY 1021 gccatgtgcgacctgctgaggaagcga 1049
DB 2992 gcatgtgtgcctcgtccggaagatgagcga 3020

```

RESULT 7
AAK19900
ID AAK19900 standard; DNA; 2293 BP.
XX

AC AAK19900;
DT 11-JUN-1999 (first entry)
XX

DE Plasmid pOC304M containing mPl promoter and Cre coding sequences.
XX

KW Mouse; mPl: promoter; site-specific recombination; Cre recombinase;
KM protamine 1; fusion gene; transgenic mouse; ss.
OS

OS Synthetic.
XX Mus musculus.
XX

PN MO9910488-A1.
XX

PD 04-MAR-1999.
XX

PF 28-AUG-1998; 98MO-US17852.
XX

PR 28-AUG-1997; 97US-0919501.
XX

PA (SALK) SALK INST BIOLOGICAL STUDIES.
XX

PI O'Gorman S, Wahl G;
XX

DR WPI: 1999-190617/16.
XX

PT New nucleic acid construct, comprising a germline-specific promoter
operably linked to a recombinase coding sequence - useful for
producing subtle, conditional or tissue-specific mutations
XX

PS Example 1; Page 48; 54pp; English.
XX

XX The present invention describes a nucleic acid construct, comprising a
germline-specific promoter operably linked to a recombinase coding
sequence. The construct can be useful for modulating agronomic
characteristics, including (a) biotic stress tolerance, increased
resistance to herbicides, pest damage, and viral, bacterial, and fungal
diseases, improvement of crop nutritional quality, reduction of
post-harvest losses, improved quantity and composition of endogenous
compounds, and production of new plant-derived compounds. The
production of alleles containing single recombinase target sites and a
desired mutation are useful for producing subtle and conditional
mutations that require generation of alleles with minimal structural
alterations. They are also useful for generating null alleles.
XX mPl-mediated nucleic acid constructs are useful for producing embryos
containing genetically lethal alleles. The recombinase constructs are
expressed at high levels in the germ line, but not to a functionally
significant degree in ES cells or embryonic or adult somatic tissue.
XX This enables modulation of recombinase target nucleic acid sequences in
the early embryo. In addition, tissue-specific and conditional
recombinase tools permit the generation of transgenic plants and
XX animals. The present sequence represents a plasmid containing mouse
protamine 1 promoter and Cre coding sequences from an example of the
present invention.
XX

Sequence 2293 BP; 555 A; 559 C; 603 G; 576 T; 0 other;

Query Match 60.6%; Score 636.4; DB 20; Length 2293;
Best Local Similarity 75.5%; Pred. No. 5.3e-85;
Matches 790; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```

QY 4 cccaagaagaagaggaagtgagcaacctgtgacccgtgacacgaacctgcccgcctg 63
DB cccaagaagaagaggaagtgagcaacctgtgacacgaacctgcccgcctg 758
QY 64 ccggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 123
DB ccggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 818
QY 124 cagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 183
DB cagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 878
QY 184 tggtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 243
DB tggtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 938
QY 244 ctgctgtacacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 303
DB ctgctgtacacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 998
QY 304 ctgagacatgctgacacgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 363
DB ctgagacatgctgacacgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1058
QY 364 ctggtgatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 423
DB ctggtgatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1118
QY 424 ctgagccttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 483
DB ctgagccttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1178
QY 484 tggcagaagacatccgacacgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 543
DB tggcagaagacatccgacacgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1238
QY 544 gccgagatgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 603
DB gccgagatgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1239
QY 604 atccacatcgccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 663
DB atccacatcgccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1358
QY 664 ctgagcgtgacacgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 723
DB ctgagcgtgacacgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1418
QY 724 aacaactacctgtctgctgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 783
DB aacaactacctgtctgctgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1419
QY 784 cagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 843
DB cagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1479
QY 844 gccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 903
DB gccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1539
QY 904 gccgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 963
DB gccgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1599
QY 964 tggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1023
DB tggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1659
QY 1024 atggtgcgcctgctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1049
DB atggtgcgcctgctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1719

```

```

RESULT 8
AAH26331
ID AAH26331 standard; DNA; 5261 BP.
XX
AC AAH26331;
XX
DT 02-OCT-2001 (first entry)
XX
DE Recombinant adenovirus delta-E1 Cre-LoxP Ad.
XX
KW Adenovirus; delta-E1 Cre-LoxP Ad; adeno-associated virus; vector;
KW gene therapy; ds.
XX
OS Chimeric - Mastadenovirus.
OS Chimeric - Human cytomegalovirus.
OS Chimeric - Rhesus macaque polyoma virus.
XX
FH Key
FH Location/Qualifiers
FT promoter
FT 51..1186
FT /tag= a
FT /note= "CMV promoter"
FT 2251..2476
FT /tag= b
FT /note= "SV40 polyA site"
FT 2476..2520
FT /tag= c
FT /function= "loxP site"
XX
PN W0200155361-A2.
XX
PD 02-AUG-2001.
XX
PE 26-JAN-2001; 2001W0-US02709.
XX
PR 26-JAN-2000; 2000US-0178536.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Hardy SF;
XX
DR WPI: 2001-483239/52.
XX
DR Producing recombinant adeno-associated virus (rAAV) vector, by stably
PT infecting eukaryotic host cell with rAAV vectors, helper adenovirus or
PT herpes virus; recombinant adenovirus or herpes vectors -
XX
PS Disclosure: Page 60-62; 63pp; English.
XX
CC The present sequence is that of recombinant adenovirus delta-E1
CC Cre-loxP Ad, in which the E1 gene of the wild-type adenovirus is
CC substituted by plasmid DNA comprising a cytomegalovirus (CMV)
CC promoter inserted immediately downstream of the CMV promoter
CC followed by a sequence from SV40 that specifies polyadenylation in
CC RNA, and a loxP site. The recombinant adenovirus is used as an
CC induction system in methods of the invention. Thus, methods
CC and compositions are provided for producing recombinant adeno
CC associated virus (rAAV) vector particles by: (a) introducing into
CC a host cell (i) AAV packaging plasmid pFlxAAV (see AAH26336), (ii)
CC a recombinant viral vector encoding plasmid, and (iii) a plasmid
CC encoding herpesvirus, cytomegalovirus or adenoviral functions, or
CC a herpesvirus, cytomegalovirus or adenovirus itself, to produce
CC a flex AAV particles and rAAV particles; and (b) introducing into a
CC second host cell (i) the rAAV particles or (a), (ii) a vector
CC that directs expression of Cre, and (iii) a vector which directs
CC expression of herpes virus, cytomegalovirus or adenovirus helper
CC functions, such that rAAV vector particles are produced. The
CC vectors are useful for in vivo or in vitro gene therapy and also
CC for in vitro recombinant protein production.
XX
SQ Sequence 5261 BP; 1309 A; 1238 C; 1329 G; 1385 T; 0 other;

```

Query Match

60.0%; Score 630; DB 22; Length 5261;

[illegible]

XX (ARTE-) ARTEMIS PHARM GMBH.
PA
XX Schwenk F;
PI
XX WPl.: 2001-441873/47.
DR P-PsDB: AAE05270.
XX
PT Using site-specific DNA recombinase domain/protein transduction domain
PT fusion proteins for inducing target gene alterations in organisms or
PS cell cultures -
XX
PS
XX Claim 17; Page 44-46; 85pp; English.

CC The present invention relates to use of fusion proteins comprising
CC a site-specific DNA recombinase domain e.g. Cre and a protein
CC transduction domain (PTD) e.g. the Human immunodeficiency virus
CC (HIV) derived TAT peptide, for preparing an agent for inducing
CC target gene alterations in a living organism or cell culture. The
CC present invention also provides a method for inducing gene of the
CC alterations in living organisms using the fusion proteins of the
CC invention. The present sequence is a DNA encoding delta
CC VP22Cre-StreptTag fusion protein. The VP22 sequence is from
CC Human spumaretrovirus (HSV).
XX

Sequence 1623 BP; 368 A; 440 C; 471 G; 344 T; 0 other:

Query Match	58.8%	Score 617.4	DB 22	Length 1623
Best Local Similarity	75.1%	Pred. No. 3.3e-82		
Matches 771	Conservative	0	Mismatches 256	Indels 0
QY 23	tgagcaacgcgtgcagcctgagacagactgcccgcctgcgcgtgagacgcacacagcg	82		
Db 542	tgcccaattactgcagccgtacacccaatttcgctccattaccggtcgaatgacaagatg	601		
QY 83	acgaagctgcgcaagaacctgatactgacatgtccgcagaccgcagagcttcacagacaca	142		
Db 602	atgaagctcgaagaacctgatactgacatgtccgcagatccgcagagcttccttgacata	661		
QY 143	ccctggagaagctgctgcgtgacgctgtgcgcgcagcttggtgcgcctgtgtcaagctgaacaac	202		
Db 662	ccctggaaaaatgctctcgtccgcgttttcgcgcgtcgttggtgcgcagctggtcgaagtgtgataaac	721		
QY 203	gcaagctgctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	262		
Db 722	ggaatggttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	781		
QY 263	gcgcgcctggtcgtgagacatccagcagaccttggtgcgcgcgcgcgcgcgcgcgcgcgcgcgc	322		
Db 782	gcgcgtctggtcagtaaaaactatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	841		
QY 323	gcagcgcgcctgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	382		
Db 842	ggtccgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	901		
QY 383	gcaagcagagacgcttgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	442		
Db 902	gaaagaaaaaacgtgatgctgcgttgagacgtgcgaacagcgcctagcgttcgaaagcactg	961		
QY 443	acttgaccagcgttggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	502		
Db 962	atttcgcacacggtctcgttccactacatggaataatgacatgctcgcgcgcgcgcgcgcgcgcgc	1021		
QY 503	tggtcctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	562		
Db 1022	tggtcctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1081		
QY 563	gcgtgaagacatacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	622		
Db 1082	gggttaagatatccacgactactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1141		
QY 623	agacccctggtgagcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	682		

[illegible]

```

Db 1310 aaagctggttagcaccgcaaggtgtagagaagcgacttagcctggtgggttaactaacctg 1369
Oy 683 tggagcgtgtagcagcgtagcgagcgagcgagcccaaacactggttctgccc 742
Db 1370 tcgagcagatgagattcgcctcctggttagctgtagatccgaataaactactgtttgccc 1429
Oy 743 gcttgcgaagaacggtgagcgcccccaagcgcaacagcagctgagcaccgggccc 802
Db 1430 ggttcagaaaaaatgtgttcgcgcgcacatctgcacacagcagctatcaactcgcgcc 1489
Oy 803 tggagggatcttcgagagccacccacgcgcctgactacgagcgcaagaagacagcgccc 862
Db 1490 tggaaaggatttttgaacacatcatctgatttagtcggtcgtcaagatgacctgtgtc 1549
Oy 863 agcgtactctgagcgagcgccgcaacgagcggtgagggcgagcgagcagctgagccc 922
Db 1550 agagctactctgctgctgtgtagacagtgccgtgctgcgagcgcgcggaataatggccc 1609
Oy 923 ggcgcggtgtagcatcccgagatcatgcaagcgcgctgtagcacaacgtgacaatcg 982
Db 1610 ggcgtgaggttcaataacaggaagatcatgcaagctgtgtgtagcacaatgttaatttg 1669
Oy 983 tgatgaactatccgcaactgtagcagcgagcgagcgagcgagcagctgctgtagg 1042
Db 1670 tcatgaactatccgcaactgtagcagcgagcgagcgagcgagcgagcgagcgagcgag 1729
Oy 1043 acggcgga 1049
Db 1730 atggcgga 1736

RESULT 13
AAV05702
ID AAV05702 standard; cDNA; 2004 BP.
XX
AC AAV05702;
XX
DT 12-MAY-1998 (first entry)
XX
DE Chimeric Cre-LDB-GR(I747T) coding sequence.
XX
KW Mutant; nuclear glucocorticoid receptor; ligand binding domain;
KW point mutation; site-directed mutagenesis; fusion protein; recombinase;
KW bacteriophage P1; gene therapy; recombination; loxp; ss.
XX
OS Chimeric - Bacteriophage P1.
OS Chimeric - Homo sapiens.
XX
FH Key
FH CDS
FT 1.2004
FT Location/Qualifiers
FT /*tag= a
FT /product= "Cre-LDB-HGR(I747T) fusion protein"
FT 1..1029
FT /*tag= b
FT /label= A
FT /note= "derived from bacteriophage P1 Cre gene"
FT 1030..1131
FT /*tag= c
FT /label= B
FT /note= "derived from C-terminus of human glucocorticoid
FT receptor D region"
FT 1132..1869
FT misc_feature
FT 1132..1869
FT /*tag= d
FT /label= C
FT /note= "derived from LDB-HGR(I747T) sequence"
FT 1876..2001
FT /*tag= e
FT /label= D
FT /note= "derived from region F of human oestrogen
FT receptor sequence"
XX
PN FR2745008-A1.
XX

```

```

PD 22-AUG-1997.
XX
PF 20-FEB-1996; 96FR-0002060.
XX
PR 20-FEB-1996; 96FR-0002060.
XX
PA (ASRE-) ASSOC DEV RECH EN GENETIQUE MOLECULAIRE.
XX
PI Brocard JB, Chambon PH, Gronemeyer H, Metzger D;
PI Nicolas JC, Roux S;
XX
DR P-PSDB; AAM44701.
XX
PT WPI; 1997-451186/42.
PT DNA encoding mutated nuclear glucocorticoid receptor - and vector
PT system for protein expression inducible by synthetic glucocorticoid
PT ligand
XX
XX
PS Claim 25; Page 54-57; 92pp; French.
XX
CC This nucleotide sequence encodes a chimeric protein comprising the Cre
CC recombinase from bacteriophage P1 linked to the ligand binding domain
CC (LBD) of the mutant nuclear glucocorticoid receptor (NGR; AAM44700). The
CC LBD has a mutation such that the activity of the receptor is induced
CC more strongly by a synthetic glucocorticoid ligand than by a natural
CC glucocorticoid ligand. The sequence encoding the fusion protein,
CC and vectors containing it, are used to treat cells either ex vivo or
CC in vitro for use in gene therapy. Vectors containing and expressing the
CC protein allow the transfer of heterologous genes to the cells' genome by
CC recombinase-directed recombination at loxp sites. By using a mutated
CC LBD, normal physiological levels of the ligand will not induce
CC recombination. This requires an increased level of a synthetic ligand.
XX
SQ Sequence 2004 BP; 548 A; 457 C; 499 G; 500 T; 0 other:

Query Match 58.8%; Score 617.4; DB 18; Length 2004;
Best Local Similarity 75.1%; Pred. No. 3,2e-82;
Matches 771; Conservative 0; Mismatches 256; Indels 0; Gaps 0:

Oy 23 tgaagcaactgtctgacgtgtgacaggaacgtgcccgtgagtgagcaccagcg 82
Db 2 tgtccaattactgacgtgtgacaggaacgtgcccgtgagtgagcaccagcg 61
Oy 83 acgaaggtgcaagaacccgtgagcaatgttccgcgagcgagcgagcgagcgagcgagc 142
Db 62 atgaaggttgcgaagaacccgtgagcaatgttccgcgagcgagcgagcgagcgagcgagc 121
Oy 143 cctggaagatgtctgtgagcggtgtgacgagcgagcgagcgagcgagcgagcgagcgagc 202
Db 122 cctggaagatgtctgtgagcggtgtgacgagcgagcgagcgagcgagcgagcgagcgagc 181
Oy 203 gcaagtgtgtcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 262
Db 182 ggaagtgtgtcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 241
Oy 263 ggcgcgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 322
Db 242 ggcgcgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 301
Oy 323 gcaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 382
Db 302 gcaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 361
Oy 383 gcaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 442
Db 362 gcaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 421
Oy 443 acttcgacaggtgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 502
Db 422 atttcgacaggtgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 481
Oy 503 tggcctcctgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 562

```

```

Db 482 tggcattctcgggagatgcttaataacaccctgttaacgtatagccgaatttcgcagatca 541
OY 563 ggcgtgaagacatcacgcgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 622
Db 542 gggtttaagataatccacgcactacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 601
OY 623 agaccctgtgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 682
Db 602 aaacgctgtgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 661
OY 683 tggagcgcgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 742
Db 662 tggagcgcgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 721
OY 743 ggcgtgacgaagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 802
Db 722 ggggtcagaataaattgtgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 781
OY 803 tggagcgcgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 862
Db 782 tggagcgcgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 841
OY 863 agcgcgtacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 922
Db 842 agagatacctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 901
OY 923 ggcgcgcgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 982
Db 902 ggcgcgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 961
OY 983 tgaatgacatcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1042
Db 962 tcatgacatcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1021
OY 1043 acgcgcga 1049
Db 1022 atggcga 1028

RESULT 14
AD09259
ID AD09259 standard; DNA; 2004 BP.
XX
AC AAD09259;
XX
DT 12-SEP-2001 (first entry)
XX
DE VP22-Cre fusion DNA.
XX
KW DNA recombinase domain; protein transduction domain; PTD;
KW gene alteration; VP22-Cre fusion protein; Human immunodeficiency virus;
KW HIV; Human spumaretrovirus; HSV; ds.
XX
OS Chimeric - Human spumaretrovirus.
OS Chimeric - Unidentified.
XX
XX
FH Key Location/Qualifiers
FT 1..2004
FT CDS /*tag= a
FT /product= "VP22-Cre fusion protein"
XX
XX
XX WO200149832-A2.
XX
XX 12-JUL-2001.
XX
XX 05-JAN-2001; 2001MO-EP00060.
XX
XX 07-JAN-2000; 2000EP-0100351.
XX
XX 10-NOV-2000; 2000EP-0124595.
XX
XX (ARTE-) ARTEMIS PHARM GMBH.
XX

```

```

PI Schwenk F;
XX
XX WPI; 2001-441873/47.
DR P-PSDB; AAE05266.
XX
XX Using site-specific DNA recombinase domain/protein transduction domain
PT fusion proteins for inducing target gene alterations in organisms or
PT cell cultures -
XX
XX Claim 17; Page 32-35; 85pp; English.
XX
XX The present invention relates to use of fusion proteins comprising
CC a site-specific DNA recombinase domain e.g. Cre and a protein
CC transduction domain (PTD) e.g. the Human immunodeficiency virus
CC (HIV) derived TAR peptide, for preparing an agent for inducing
CC target gene alterations in a living organism or cell culture. The
CC present invention also provides a method for inducing gene
CC alterations in living organisms using the fusion proteins of the
CC invention. The present sequence is a DNA encoding VP22-Cre fusion
CC protein. The VP22 sequence is from Human spumaretrovirus (HSV).
XX
XX Sequence 2004 BP; 419 A; 608 C; 592 G; 385 T; 0 other;
SQ

Query Match 58.8%; Score 617.4; DB 22; Length 2004;
Best Local Similarity 75.1%; Pred. No. 3.2e-62;
Matches 771; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

OY 23 tgaagcaactgtgacacgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 82
Db 974 tgcacaattcagcgtacacgaacaaatttgcctgattacacggttgcagacagtg 1033
OY 83 acgaagtcgacgaacacgtgacacatgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 142
Db 1024 atgaggttcgacgaacacgtgacacatgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1093
OY 143 cctggaagatctgctgagcgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 202
Db 1094 cctggaagatctgctgagcgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1153
OY 203 gcaagtggtcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 262
Db 1154 ggaatgtgtcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1213
OY 263 ggcgcgtgcgcgtgacacacacacacacacacacacacacacacacacacacacacacacac 322
Db 1214 ggcgcgtgcgcgtgacacacacacacacacacacacacacacacacacacacacacacacac 1273
OY 323 gcaagcgcgtcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 382
Db 1274 ggcgcgcgtcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1333
OY 383 gcaaggaagacgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 442
Db 1334 gaaaggaagacgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1393
OY 443 acttcgacacaggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 502
Db 1394 atttcgacacaggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1453
OY 503 tggccttcctgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 562
Db 1454 tggcattctcgggagatgcttaataacaccctgttaacgtatagccgaatttcgcagatca 1513
OY 563 ggcgtgaagacatcacgcgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 622
Db 1514 gggtttaagataatccacgcactacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1573
OY 623 agaccctgtgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 682
Db 1574 aaacgctgtgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1633
OY 683 tggagcgcgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 742

```

```

Db      1634  tcgagcagatggattccgtcctcgtggtatgctgatatccgatatacctaactccgttttgc 1693
Oy      743   gcgtgcaagaacgagcgcgcgcccccacgacgacgacgacgtgagcaaccgggccc 802
Db      1694  gggctcagaaaatggtgtgttcgcgcacatctgcacacgacgacgactatcaactcgccc 1753
Oy      803   tggagggcactctcgagggcaccaccccgctgatacggcgccgaagaaacgacggcc 862
Db      1754  tggaaaggatttttgaagaactatcatgatttaacggcgctaaagatgactctgtgc 1813
Oy      863   aacgctactcgtcgagcgacgacgacgacgacgacgacgacgacgacgacgacgac 922
Db      1814  agagatactcgtcgctgtgtgacacagtgcccgctcgagcgacgacgacgacgacgac 1873
Oy      923   gcgcgcgctgagcgcaccccgacatcatgacgacgacgacgacgacgacgacgacgac 982
Db      1874  ggcgtgaggtttcattcaataccgagatcatgacgacgacgacgacgacgacgacgacgac 1933
Oy      983   tgaataactacatccgcaacgtgagacgacgacgacgacgacgacgacgacgacgacgac 1042
Db      1934  tcatgaactataccgttaactcgtgatactgataaagcgagcgacgacgacgacgacgacgac 1993
Oy      1043  acgagcga 1049
Db      1994  atggcgga 2000

```

RESULT 15

AAD09268
ID AAD09268 standard; DNA; 2055 BP.

AC AAD09268;

DT 12-SEP-2001 (first entry)

DE VP22Crestreptag fusion DNA.

XX DNA recombinase domain: protein transduction domain; PRD;

KM VP22Crestreptag fusion protein; Human Immunodeficiency Virus; HIV;

KM gene alteration; Human spumaretrovirus; HSV; ds.

OS Chimeric - Human spumaretrovirus.

OS Chimeric - Unidentified.

FH Key Location/Qualifiers

FT 1..2052

FT CDS /tag= a /Product= "HSV VP22Crestreptag fusion protein"

PN WO200149832-A2.

PD 12-JUL-2001.

PF 05-JAN-2001: 2001WO-EP00060.

PR 07-JAN-2000: 2000EP-0100351.

PR 10-NOV-2000: 2000EP-0124595.

XX (ARTE-) ARTEMIS PHARM GMBH.

PI Schwenk F;

DR WPI: 2001-441873/47.

DR P-PSDB: AAE05273.

XX Using site-specific DNA recombinase domain/protein transduction domain

PT fusion proteins for inducing target gene alterations in organisms or

XX cell cultures -

PS Claim 17: Page 55-58; 85pp; English.

XX The present invention relates to use of fusion proteins comprising

CC a site-specific DNA recombinase domain e.g., Cre and a protein
CC transduction domain (PRD) e.g., the Human immunodeficiency virus
CC (HIV) derived RAP peptide, for preparing an agent for inducing
CC target gene alterations in a living organism or cell culture. The
CC present invention also provides a method for inducing gene
CC alterations in living organisms using the fusion proteins of the
CC invention. The present sequence is a DNA encoding VP22Crestreptag
CC fusion protein. The VP22 sequence is from Human spumaretrovirus
CC (HSV).

SO Sequence 2055 BP; 427 A; 619 C; 609 G; 400 T; 0 other;

Query Match 58.8%; Score 617.4; DB 22; Length 2055;
Best Local Similarity 75.1%; Pred. No. 3,2e-82;
Matches 771; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```

Oy      23   tggagcaaccgtctgacgcgtgcacagaacctgcgcgcctgcgttgagacgacacagcg 82
Db      974   tgcataattactgacacgttacccaatattgcctgcatatccgcgtcgtatgacagagtg 1033
Oy      83   acgagtgccgcaagaacactgatatgacatgttcgcgcacgacgacgacgacgacgacgac 142
Db      1034  atgaggttcgcaagaacactgatatgacatgttcgcgcacgacgacgacgacgacgacgac 1093
Oy      143   cctggaagatgctgctgagcgtgttcgcgcacgctggtggtgacgacgacgacgacgac 202
Db      1094  cctggaanaatgcttctgtcctgttcgttcgcgcgtggtggtgacgacgacgacgacgac 1153
Oy      203   gcaagtgttcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 262
Db      1154  ggaatggtttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1213
Oy      263   gcgcgttcgcgcgttgagacacatccagcagcactgtggtgcacgtgacatcgtccgcgc 322
Db      1214  gcggttcgcgcgttgagacacatccagcagcactgtggtgcacgtgacatcgtccgcgc 1273
Oy      323   gcgcgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 382
Db      1274  ggtccgcgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1333
Oy      383   gcaagagaacgltgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 442
Db      1334  gaaagaaaacgltgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1393
Oy      443   acttcgacacaggttgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 502
Db      1394  atttcgacacaggttgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1453
Oy      503   tggccttcctggcactcgcctcaacaaccctgtcgcgcacgcgcgcgcgcgcgcgcgcgc 562
Db      1454  tggcattctctgggagattgctatacaaccctgttcgcgcacgcgcgcgcgcgcgcgcgcgc 1513
Oy      563   gcgtgaaagacatcagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 622
Db      1514  gggttaaaagatactcagcgtactcagcgttggagaaatgttaatacattgcaagaagca 1573
Oy      623   agaccctgtgagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 682
Db      1574  aaacgctgttagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1633
Oy      683   tggagcgtgatacagcgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1693
Db      1634  tcgagcgaatgattctcgtctctgtgtgagtgatgatacgaataactactcgttttgcgc 1693
Oy      743   gcgtgcaagaacgagcgcgcccccacgacgacgacgacgacgacgacgacgacgacgacgac 802
Db      1694  ggcctcagaaaatggtgtgttcgcgcacatctgcacacgacgacgacgacgacgacgacgac 1753
Oy      803   tggagggcactctcgagggcaccaccccgctgatacggcgccgaagaaacgacgacgacgac 862
Db      1754  tggaaaggatttttgaagaacatcatcgtattgatttaacggcgctgaagatgactctgtgc 1813

```

```

QY 863 agcgctacctgagcctgagcgcgacagcgcccgctggtggcgccgcgcgcgacatggccc 922
    ||| ||||| ||||| ||| ||||| ||||| ||| ||||| ||||| ||||| |||||
Db 1814 agagataacctggcctggtctggaacacagtgccggtgtcggaagcgcgagatattggccc 1873
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 923 ggcgcgcgcgtgagacatcccgagatcatgagcgccgcgcgtggaaccaatgaaatcg 982
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1874 ggcgtggagttcaataaccggaatcatgcaagctgtgtgctggaaccaatgtaaatgtg 1933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 983 tgatgaactaatccgaacctggacagcgagacggcgccaatgtgtgcctgtggaag 1042
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1934 tcaatgaactaatccgaacctggacagcgagacggcgccaatgtgtgcctgtggaag 1993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1043 acggcgga 1049
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1994 atggcgga 2000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: May 10, 2002, 02:06:46
Job time: 4323 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2002, 00:50:43 ; Search time 71.81 Seconds
(without alignments)
3591.633 Million cell updates/sec

Title: US-09-662-128A-1

Perfect score: 1050
Sequence: 1 atccccagaagaagagagaa.....gcctgtgtgagagcgcgac 1050

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/prodata/2/lna/5A_COMB.seq: *
2: /cgn2_6/prodata/2/lna/5B_COMB.seq: *
3: /cgn2_6/prodata/2/lna/6A_COMB.seq: *
4: /cgn2_6/prodata/2/lna/6B_COMB.seq: *
5: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/2/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
1	755	71.9	2346	4	US-09-193-503B-5	Sequence 5, Appli
2	754.4	71.8	1032	4	US-09-193-503B-2	Sequence 2, Appli
3	617.4	58.8	1740	2	US-08-864-224-10	Sequence 10, Appli
4	617.4	58.8	2346	4	US-09-193-503B-4	Sequence 4, Appli
5	617.4	58.8	2346	4	US-09-193-503B-7	Sequence 7, Appli
6	241.2	23.0	699	2	US-08-735-609-11	Sequence 11, Appli
7	241.2	23.0	699	2	US-08-735-609-11	Sequence 11, Appli
8	241.2	23.0	699	2	US-09-315-372-11	Sequence 11, Appli
9	241.2	23.0	699	2	US-09-244-752-11	Sequence 11, Appli
10	241.2	23.0	699	2	US-09-244-752-11	Sequence 11, Appli
11	241.2	23.0	699	2	US-09-244-752-11	Sequence 11, Appli
12	88.2	8.4	3624	1	US-07-951-715A-6	Sequence 6, Appli
13	88.2	8.4	3624	1	US-07-951-715A-6	Sequence 6, Appli
14	88.2	8.4	3624	1	US-08-459-448A-6	Sequence 6, Appli
15	88.2	8.4	3624	1	US-08-459-595A-6	Sequence 6, Appli
16	88.2	8.4	3624	1	US-08-459-504B-6	Sequence 6, Appli
17	88.2	8.4	3624	1	US-08-459-504B-6	Sequence 6, Appli
18	88.2	8.4	3624	1	US-09-053-549-7	Sequence 7, Appli
19	88.2	8.4	3624	1	US-09-547-422-6	Sequence 6, Appli
20	86.6	8.2	3468	1	US-07-951-715A-2	Sequence 2, Appli
21	86.6	8.2	3468	1	US-08-459-448A-2	Sequence 2, Appli
22	86.6	8.2	3468	1	US-08-459-595A-2	Sequence 2, Appli
23	86.6	8.2	3468	1	US-08-459-504B-2	Sequence 2, Appli
24	86.6	8.2	3468	1	US-08-459-444-2	Sequence 2, Appli
25	86.6	8.2	3468	1	US-09-053-549-3	Sequence 3, Appli
26	85.6	8.2	975	4	US-09-547-422-2	Sequence 2, Appli
27	85.2	8.1	1965	4	US-09-178-252-26	Sequence 26, Appli

28	82.2	7.8	8654	3	US-09-053-549-1	Sequence 1, Appli
29	81.2	7.7	3468	1	US-07-951-715A-4	Sequence 4, Appli
30	81.2	7.7	3468	2	US-08-459-448A-4	Sequence 4, Appli
31	81.2	7.7	3468	3	US-08-459-595A-4	Sequence 4, Appli
32	81.2	7.7	3468	3	US-08-459-504B-4	Sequence 4, Appli
33	81.2	7.7	3468	3	US-08-459-444-4	Sequence 4, Appli
34	81.2	7.7	3468	3	US-09-053-549-5	Sequence 5, Appli
35	81.2	7.7	3468	4	US-09-547-422-4	Sequence 4, Appli
36	80.8	7.7	975	4	US-09-365-150-3	Sequence 4, Appli
37	77.6	7.4	2634	1	US-08-196-218-31	Sequence 31, Appli
38	77.6	7.4	2634	1	US-08-681-953-31	Sequence 31, Appli
39	77.6	7.4	2943	1	US-08-042-747A-7	Sequence 7, Appli
40	76.6	7.3	12588	2	US-08-387-942C-1	Sequence 1, Appli
41	76.2	7.3	1140	3	US-09-023-173-4	Sequence 4, Appli
42	76	7.2	1235	1	US-08-457-797A-11	Sequence 11, Appli
43	76	7.2	1235	1	US-08-812-025-11	Sequence 11, Appli
44	76	7.2	1235	4	US-09-138-873A-11	Sequence 11, Appli
45	76	7.2	2219	3	US-08-510-646B-17	Sequence 17, Appli

ALIGNMENTS

```
RESULT 1
US-09-193-503B-5
; Sequence 5, Application US/09193503B
; Patent No. 6262341
; GENERAL INFORMATION:
; APPLICANT: Baszczyński, Christopher L.
; APPLICANT: Lyszniak, Leszek A.
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Guan, Xueni
; APPLICANT: Rao, Guru
; APPLICANT: Tagliani, Laura A.
; TITLE OF INVENTION: A No. 6262341e1 Method For The Integration Of Foreign DNA In
; FILE REFERENCE: Eukaryotic Genomes
; CURRENT FILING DATE: 1998-11-17
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/099,435
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: 60/056,627
; PRIOR FILING DATE: 1997-11-18
; PRIOR APPLICATION NUMBER: 60/065,613
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: sequence
; OTHER INFORMATION: encoding mcre:FLPM, Cre from Bacteriophage P1 and
; OTHER INFORMATION: FLP from Saccharomyces, both maize preferred
; NAME/KEY: CDS
; LOCATION: (1)..(2346)
; US-09-193-503B-5
```

```
Query Match 71.9%: Score 755; DB 4; Length 2346;
Best Local Similarity 83.4%: Pred. No. 4,6e-121;
Matches 857; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
QY 23 tgagcaactcgtacacgtgcacagaacctgcgcgcctcgtgcgcgttcagcgccacagcg 82
Db 2 tttcctacactcgtacacgttcacacgaaccttcgcgccttcacgtgcgcgcgcgcgcg 61
QY 83 acgagctcgcgaagaacctgtatgacatgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 142
Db 62 atgaagtcaggaagaacctcattgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 121
```


D b	602	a g a g c t c g t t t c c a a c c g c a g c g t t c g a a a a g g c c t c a g c t c t c g a g t t g a c c a a g c t c g	661
Q y	683	t g a g a c g c t g a t c a t c a g c g t g a g c g c g t g g c c g a g a c a c c o c a a c a a c t a c t c g t t c t g c c	742
D b	662	t c g a a a c c t g a t c t c c g t c t c c g t c g t c g a g a c c o c a a c a a c a a c t a c t c t c t t c g c c	721
Q y	743	g c g t g c g c a a g a a c g u g u g t g c c g c c c a g c g c a c a g c c a a g c t a g t a g a c c c u g u g c c c	802
D b	722	g c g a c c a g c a a a a c g u g u g t g a g c t g a c c c t e a g c t c g c c a c c a g c a c a t c a g c a a g a g g c c t	781
Q y	803	t g a g a g g a c a t t c g a g a g c a c a c c a c c a c g c t g a t a t a g t a g c g c a a g a g a g a g a c a g c c	862
D b	782	t g a g a a g t a t t t c g a g a g c a c a c c a c c g c t c a g t a c a a g u g c g a a g a g a t g a c a g u g t c	841
Q y	863	a a g c g t a c t c t g c c c t t g a g a c g g c a a g a c a g c c g c c g t t g u g c g c g c c a c c g a c a t g u g c c	922
D b	842	a a c g c t e a c t c g a t g t t c c g g a c a t c c g c c g c g t t g a g a c t g a g a g a c t a g c a t g c c	901
Q y	923	g c g c a g c g t g a g a c a t c c c c g a g a t a t a t c a g a c c g c g c t c g a g a c a a c t g t a a c a t c g	982
D b	902	g c g c a g t g t t t c c a t c c c c g a a t a t a t a t c a t a t c a g a c g u g u g t g a t g a g c u g a a c g t g a a c a t t g	961
Q y	983	t g a t a a a c a t a c a t c c g c a a c t t g a a a g u g a g a c c g c a t g t g c a t g t c g c t c g t c g a g a g	1042
D b	962	t c a t a a c t a c a t t c g c a a c c t t g a a a g a g a g a c g u g c g a t g t t c g c t c c t c g a a g	1021
Q y	1043	a c g g c g a c	1050
D b	1022	a t g t g a c	1029

```

RESULT 3
US-08-864-224-10
: Sequence 10, Application US/08864224
: Patent No. 5851808
:
: GENERAL INFORMATION:
: APPLICANT: Elledge, Stephen J.
: APPLICANT: Liu, Qinghua
: TITLE OF INVENTION: Rapid Subcloning Using Site-Specific
: TITLE OF INVENTION: Recombination
: NUMBER OF SEQUENCES: 18
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Medlen & Carroll, LLP
: STREET: 220 Montgomery Street, Suite 2200
: CITY: San Francisco
: STATE: California
: COUNTRY: United States of America
: ZIP: 94104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/864,224
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Ingolia, Diane E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: BOM-02681
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1740 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "DNA"
: FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 1..1737
US-08-864-224-10

```

Query Match	58.8%	Score 617.4;	DB 2;	Length 1740;
Best Local Similarity	75.1%;	Pred. No. 1.3e-97;		
Matches 771; Conservative	0;	Mismatches 256;	Indels 0;	Gaps 0

[illegible]

[illegible]

```

; NUMBER OF SEQ ID NOS: 11
;
; SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 7
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: sequence
; OTHER INFORMATION: encoding a Cre:FLP polypeptide, Cre from
; OTHER INFORMATION: Bacteriophage P1 and FLP from Saccharomyces
US-09-193-503B-7

```

```

Query Match          58.8%; Score 617.4; DB 4; Length 2346;
Best Local Similarity 75.1%; Pred. No. 1.3e-97;
Matches 771; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```

```

QY 23 tgaagcaactgtgacgttcgacagcaacccgctccgctgagccgacccagcg 82
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 tggccaattactgacgttaccacaaatttgcctgcattaccggtcgaatgcaagatg 61

QY 83 acgaagtgccgaagaacctgatatgacatgttccgcagccagcccttcagcgacaca 142
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 atgaggttcgcaagaacctgatatgacatgttccgagatccgcagcgctttctgacata 121

QY 143 ccttgaagaatgctctgtagcggtgtgcgcagctgagccgctgtgtgcaagctgaacaac 202
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 ccttgaagaatgctctgtagcggtgtgcgcagctgagccgctgtgtgcaagctgaacaac 181

QY 203 gcaagtgttcccgccgagccgagcgatgagcgactactgtctgtacttcgacggccc 262
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 ggaatgttcccgccgagccgagcgatgagcgactactgtctgtacttcgacggcgcc 241

QY 263 gcgagctgtgcgttgaagacacatccagcagcccttgagcgacgtgaacatgtgcacgcc 322
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 gcggtctgagcagtaaaactatccagcaacatttgagcgacgtgaacatgttcacgtc 301

QY 323 gcaagcgctgtcccgcccgccagcgacacacacgcgtgtagctgtgtgacgcgcacatcc 382
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 ggtccggtgtgcacgacacaaatgacacgaatgtctgttcaactgtgtatgcgcgcagatcc 361

QY 383 gcaaggaagaacgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 442
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 gaaagaagaacgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 421

QY 443 acttcgacacaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 502
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 atttcgacacaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 481

QY 503 tggccttctgtgacatgcctacacacacgcgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgc 562
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 tggcattctgtggaattgctataacacccctgttactgatagcgaaattgcccaggaatca 541

QY 563 ggcgtgaagacatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 622
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 542 ggcgttaagaatatacctcagctactgcgcgttggagaaatgttatccatatttgcagaagaca 601

QY 623 agaccctgtgtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 682
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 602 aaacgttgtttagcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 661

QY 683 tggagcgctgtgacgcgcgtgtgagcgcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 742
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 662 tggagcgagatgacgttccgtctcgtgtgtagctgtagatccgaataatactactgttttcc 721

QY 743 gctgtgcgaagaagacgcgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 802
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 722 ggcgtcagaanaaaatgtgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 781

QY 803 tggagggcatctctgagcgacacccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 862
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 782 tggagggatcttctgaagaacactatcgatttactgagcgctaaaggaatgatctgtctc 841

QY 863 aagcgtacctgagcgtgtgagcgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 922
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 842 agagataccctggctgtgtgacacagatgcccgtgtcggagcgccgcgagataatgccc 901
QY 923 ggcgcgcgtgtgagcatcccgagatcatcagagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 982
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 902 ggccttgagtttcaatcccgagatcatcagagctgtgtgtgtgacccaatgtaataatg 961

QY 983 tgaatatacatcccgacacactgtgacagcagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1042
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 962 tcatgaactataatccgtaacctgtgatagtgaacacggcgcaatggtgcgcctgtcgtgaag 1021

QY 1043 acggcgca 1049
   || ||| |||
Db 1022 atggcgca 1028

```

```

RESULT 6
US-09-193-503B-8
; Sequence 8, Application US/09193503B
; Patent No. 6262341
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Christopher L.
; APPLICANT: Lyszniak, Leszek A.
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Rao, Xueni
; APPLICANT: Regliant, Laura A.
; TITLE OF INVENTION: A No. 6262341el Method For The Integration Of Foreign DNA Int
; FILE REFERENCE: 5718-66 (amended listing)
; CURRENT APPLICATION NUMBER: US/09/193,503B
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/099,435
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: 60/056,627
; PRIOR FILING DATE: 1997-11-18
; PRIOR APPLICATION NUMBER: 60/065,613
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: encoding a FlpM-Cre polypeptide, Flp from
; OTHER INFORMATION: Saccharomyces (maize preferred codons), and Cre
; NAME/KEY: CDS
; LOCATION: (1)..(2346)
US-09-193-503B-8

```

```

Query Match          58.8%; Score 617.4; DB 4; Length 2346;
Best Local Similarity 75.1%; Pred. No. 1.3e-97;
Matches 771; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```

```

QY 23 tgaagcaactgtgacgttcgacagcaacccgctccgctgagccgacccagcg 82
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1316 tggccaattactgacgttaccacaaatttgcctgcattaccggtcgaatgcaagatg 1375

QY 83 acgaagtgccgaagaacctgatatgacatgttccgcagccagcccttcagcgacaca 142
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1376 atgaggttcgcaagaacctgatatgacatgttccgagatccgcagcgctttctgacata 1435

QY 143 ccttgaagaatgctctgtagcggtgtgcgcagctgagccgctgtgtgcaagctgaacaac 202
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1436 ccttgaagaatgctctgtagcggtgtgcgcagctgagccgctgtgtgcaagctgaacaac 1495

QY 203 gcaagtgttcccgccgagccgagcgatgagcgactactgtctgtacttcgacggccc 262
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1496 ggaatgttcccgccgagacccggaagatgttcgcgataatcttcatatcttcagcgcc 1555

```



```

1 TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
2 NUMBER OF SEQUENCES: 15
3 CORRESPONDENCE ADDRESS:
4 ADDRESSEE: Medien & Carroll, LLP
5 STREET: 220 Montgomery Street, Suite 2200
6 CITY: San Francisco
7 STATE: California
8 COUNTRY: United States of America
9 ZIP: 94104
10
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: Floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: Patent Release #1.0, Version #1.30
16
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/735,609
19 FILING DATE: 23-Oct-1996
20 CLASSIFICATION: <unknown>
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Ingolia, Diane E.
23 REGISTRATION NUMBER: 40,027
24 REFERENCE/DOCKET NUMBER: UM-02484
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: (415) 705-8410
27 TELEFAX: (415) 397-8338
28
29 INFORMATION FOR SEQ ID NO: 11:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 699 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35
36 MOLECULE TYPE: other nucleic acid
37 DESCRIPTION: /desc = "DNA"
38
39 US-08-735-609-11
40
41 Query Match 23.0%, Score 241.2; DB 2; Length 699;
42 Best Local Similarity 77.0%; Pred. 1.6e-33;
43 Matches 294; Conservative 0; Mismatches 88; Indels 0; Gaps
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
95
```

```

; Patent No 6057158
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amaltilano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,372
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/735,609
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-09-315-372-11

Query Match          23.0%; Score 241.2; DB 3; Length 699;
Best Local Similarity 77.0%; Pred. No.1,6e-33;
Matches 294; Conservative 0; Mismatches 88; Indels 0; Gaps 0.

OY      1 atgcccaagaagaaaggatgacgaacctgtgcacgctgaccagaacctggccgccc 60
DB      318 ATGCCCAAGAAAGGAGTGTCTCAATTTACTGACCGCTACACCATAAATTGGCTTGA 377
        |||||||
OY      61 ctgcgcgttgaaagcaccaccccgagcgaagatgctgacgaagaacttgaagaagtccggagc 120
DB      378 TTACCCGGTCGATGCACAACGATGATAGAAGTTCCGACAGAACCTGATGGACATGTTCAAGCAT 437
        |||||
OY      121 cgcacaagccttcagcagcagcacacacctggaagaatgctgtcgtgtagcgtgtgcgcagcctggagcc 180
DB      438 CGCACAAGCGCTTTCTTGAGACATACCTTGGAATAATGCTTCTGTCGTTTCCGGTGGTGGGG 497
        |||||
OY      181 gctctgtgtaaacgtcaacaacccgaagaatggtgtccccgcgcgagcccgaggaagctgtcgagac 240
DB      498 GCATGGTGCMAATTAATAATACCGGAATAGCTTCCCGCAGAACCTGAAGATGTTGGCGAT 557
        |||||
OY      241 taacctgcgttacctgcagagcccgcgagcctgagcctggaagaacacatacaagcactggagcc 300
DB      558 TATCTTCAATATACCTTCAAGCGCGGGGCTCTCGCAGTAATAAATCAATTCACACACATTTTGGGC 617
        |||||
OY      301 caagctgaacaatgctgcacccgcgcgaagcggcctgtgcccgccccaagcagacaagcgcgtg 360
DB      618 CAGCTAACATGCTTCATCAGTCGTGGTCCGGGGCTGCCAGCACCAAGTAGACAGCAATGCTGTT 677
        |||||

```

181 g c c t t g t c a a g c t g a a c a c c g c a a g t g t t c c c c g c c a g c c c g a g a c t g c g c a c 240

61 ctaccctatgacgcccaccacgacgagatgcgcacaagaacctgatacgacatgttccgcgac 120

RESULT 13
US-08-459-448A-6
; Sequence 6, Application US/08459448A
; Patent No. 5859336
; GENERAL INFORMATION:

Query Match	8.48;	Score 88.2;	DB 2;	Length 3624;
Best Local Similarity	47.0%;	Pred. No. 1.9e-07;		
Matches 310;	Conservative 0;	Mismatches 343;	Indels 6;	Gaps

RESULT 14
 US-08-459-595A-6
 ; Sequence 6, Application US/08459595A
 ; Patent No. 6018104
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Koziel, Michael G.
 ; APPLICANT: Desai, Nalini M.
 ; APPLICANT: Lewis, Kelly S.
 ; APPLICANT: Kramer, Vance C.
 ; APPLICANT: Warren, Gregory W.
 ; APPLICANT: Evola, Stephen V.
 ; APPLICANT: Crossland, Lyle D.
 ; APPLICANT: Wright, Martha S.
 ; APPLICANT: Merlin, Ellis J.
 ; APPLICANT: Lannis, Karen L.
 ; APPLICANT: Rochstein, Steven J.
 ; APPLICANT: Bowman, Cindy G.
 ; APPLICANT: Dawson, John L.
 ; APPLICANT: Dunder, Erik M.
 ; APPLICANT: Pace, Gary M.
 ; APPLICANT: Suttie, Janet L.
 ; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
 ; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
 ; NUMBER OF SEQUENCES: 94
 ;
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NO. 6018104artis Corporation
 ; STREET: Patent & Trademark Dept., 520 White Plains
 ; STREET: Rd., POB 2005
 ; CITY: Tarrytown
 ; STATE: New York


```

1 INFORMATION FOR SEQ ID NO: 6:
2
3 SEQUENCE CHARACTERISTICS:
4
5     LENGTH: 3624 base pairs
6
7     TYPE: nucleic acid
8
9     STRANDEDNESS: single
10
11    TOPOLOGY: linear
12
13    MOLECULE TYPE: other nucleic acid
14
15    DESCRIPTION: /desc = "Synthetic DNA"
16
17    HYPOTHEICAL: NO
18
19    FEATURE:
20
21        NAME/KEY: CDS
22
23        LOCATION: 1..3621
24
25        OTHER INFORMATION:
26
27        OTHER INFORMATION:
28
29        OTHER INFORMATION:
30
31        /product= "Full-length, maize
32        optimized cytb"
33        /note= "Disclosed in Figure 6."
34
35 JS-08-459-504B-6

```

[illegible]

Search completed: May 10, 2002, 02:02:12
Job time: 4289 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2002, 00:48:43 ; Search time 1779.62 Seconds
(without alignments)
7963.387 Million cell updates/sec

Title: US-09-662-128a-1

Perfect score: 1050
Sequence: 1 atgcccaagaagaagagagaa.....gcctctgagagagcgac 1050

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_estbm:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gD_estl:*
10: gD_estl2:*
11: gD_hc:*
12: gD_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	390	37.1	688	10	B1651236	B1651236 603298328
2	231	22.0	431	9	AW211323	AW211323 uc079f01.y
3	181.2	17.3	563	10	B1650191	B1650191 603296283
4	158.8	15.1	340	9	AW212478	AW212478 uc089c09.x
5	150.6	14.3	338	9	AW211972	AW211972 uc079f01.x
6	128.2	12.2	343	9	AW210918	AW210918 uc089c09.y
7	86	8.2	836	10	B1556329	B1556329 603237448
8	79.8	7.6	288	12	FR0013439	FR0013439 F. rubripes
9	75.6	7.2	925	12	CNS0091P	AL0054689 F. rubripes
10	74.4	7.1	852	10	B1949774	AL0053013 Drosophila
11	74.4	7.1	854	10	BE261766	B1949774 HVSME1001
12	72.2	6.9	1464	3	BE636746	BE261766 HV_CEA0006X
13	72	6.9	949	12	CNS031R8	BE636746 rockefeller
14	69.8	6.6	616	12	CNS03VHV	AL223901 Tetradon
15	69	6.6	935	12	CNS006XK	AL652934 Tetradon
16	68.2	6.5	637	10	BE251052	AL066051 Drosophila
17	67	6.4	692	10	BM134842	BE251052 EST418309
						BM134842 WHE0452_G

18	67	6.4	759	10 BF662500	BF662500 963035F01
19	66.8	6.4	559	12 AZ661219	AZ661219 1M0539M10
20	66.8	6.4	610	10 BM324963	BM324963 PTC1_38_A
21	66.8	6.4	1183	10 BE636683	BE636683 rockefell
22	66.6	6.3	654	6 B1140552	B1140552 IP1_50_H0
23	65.8	6.3	622	9 AV927327	AV927327 AV927327
24	65.8	6.3	683	10 B1546421	B1546421 603188722
25	65.6	6.2	702	10 B1836679	B1836679 603089610
26	65.4	6.2	1101	12 CENS017SY	AL108460 Drosophila
27	65.2	6.2	1177	10 BE636682	BE636682 rockefell
28	65	6.2	760	10 BR266871	BR266871 HV_CEA001
29	65	6.2	1453	10 BM320858	BM320858 turnast_1.1
30	64.8	6.2	503	10 BE404787	BE404787 WHE1205_E
31	64.8	6.2	964	10 BG343549	BG343549 HVSME000
32	64.6	6.2	607	10 BG39506	BG39506 EML_81_G0
33	64.6	6.2	1239	12 AG047572	AG047572 Pan trogl
34	64.4	6.1	591	10 B1718551	B1718551 1031028E0
35	64.4	6.1	857	10 BP262005	BP262005 HV_CEA000
36	64.2	6.1	485	9 AW161386	AW161386 au81b07.y
37	64.2	6.1	674	10 B1955291	B1955291 HVSME002
38	63.8	6.1	824	10 B1953697	B1953697 HVSME001
39	63.4	6.0	598	9 AW324623	AW324623 707036B03
40	63.4	6.0	710	9 AU163439	AU163439 AU163439
41	63.4	6.0	819	10 BG309572	BG309572 HVSME001
42	63.2	6.0	598	10 BE518410	BE518410 WHE0815_E
43	63.2	6.0	632	10 B1959076	B1959076 HVSME001
44	63.2	6.0	1137	10 BG809979	BG809979 mgc002xd
45	63.2	6.0	1452	12 AG032979	AG032979 Pan trogl

ALIGNMENTS

RESULT 1
B1651236
LOCUS 688 bp mRNA linear EST 12-SEP-2001
DEFINITION 603298328F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5338806 5',
B1651236
ACCESSION B1651236
VERSION B1651236.1 GI:15565472
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS 1 (bases 1 to 688)
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1860 row: h column: 07
High quality sequence stop: 688.
Location/Qualifiers
1. 688
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5338806"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site: 1; Note:
Site 2: Salt; Cloned unidirectionally; Primer: Oligo dt.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference

FEATURES

source

LOCUS B1650191 563 bp mRNA linear EST 12-SEP-2001
 DEFINITION 603296283f1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5336853 5',
 mRNA sequence.
 ACCESSION B1650191
 VERSION B1650191.1 GI:15564427
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Mus.
 REFERENCE 1 (bases 1 to 563)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1855 row: f column: 22
 High quality sequence stop: 563.
 Location/Qualifiers
 1..563
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:5336853"
 /clone_1lb="NIH_CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; NotI;
 Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert 2 kb. Library constructed by Life
 Technologies, catalog #12017-018. Investigators providing
 samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
 (1999). Note: this is a NCI CGAP Library."
 BASE COUNT 143 a 133 c 149 g 138 t
 ORIGIN
 Query Match 17.3%; Score 181.2; DB 10; Length 563;
 Best Local Similarity 76.6%; Pred. No. 6.5e-15;
 Matches 222; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 760 gtggcgcccccgccagccagccagctgagcccgccctgtgagcattccgag 819
 Db 1 GTTCCCGCGGCATCTGCGACACGACCTATCAACTGCGCCCTGGAGGATTTTGA 60
 QY 820 gccaccacgcctgatactagcgcccaagagcagcagcgccgctacccgctgag 879
 Db 61 GCAATCATGATGATTGATTACGGCGCTAAGGATGACATCGGTGAGATACCTGCGCTGG 120
 QY 880 agcgagcagcgcccgcggtggcgcccgccgagcattgcccgcggtgagcattc 939
 Db 121 TCTGGACACAGTCCCGTGTGCGGCGCGGAGATATGCGCCGCTGAGATTCAATA 180
 QY 940 cccgagatcatgacgagcgcgctgagccaacgtgaacatcgtgtgatacaccgcg 999
 Db 181 CCGGAGATCATGCAACCTGGTGTGACCAATGTAATATTTGTCTGTAAGATATATCCGT 240
 QY 1000 aacctgacagcgagcagcgccatggtgctgctgtgagagcagcga 1049
 Db 241 AACCTGATATGTAACAGGGGCAATGTCGCGCTCTGGAAGATGGCGCA 290
 RESULT 4
 LOCUS AW212478 340 bp mRNA linear EST 03-DEC-1999
 DEFINITION u089c09.x1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:2649712 3',

similar to SW:RECR_BPPI P06956 RECOMBINASE CRE. [1] ;, mRNA
 sequence.
 ACCESSION AW212478
 VERSION AW212478.1 GI:6518565
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 340)
 AUTHORS NCI-CGAP http://www.nci.nih.gov/ncicgap/
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other ESTs: u089c09.y1
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 MGI:1030164
 Seq primer: -40UP from G1bco
 High quality sequence stop: 158.
 Location/Qualifiers
 1..340
 /organism="Mus musculus"
 /strain="129 - C57/B6 - FVBN"
 /db_xref="taxon:10090"
 /clone="IMAGE:2649712"
 /clone_1lb="NCI_CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /dev_stage="10 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; SalI;
 Site:2; NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies, Inc. Investigators
 providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
 Reference for transgenic model: Xu et al., Nature Genetics
 22, 37-43 (1999)."
 BASE COUNT 75 a 100 c 79 g 86 t
 ORIGIN
 Query Match 15.1%; Score 158.8; DB 9; Length 340;
 Best Local Similarity 76.0%; Pred. No. 5.4e-12;
 Matches 196; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
 QY 792 caccggcccttgaggggcatctcgaggccaccacgctgatactacgcccaga 851
 Db 293 CAATCGCGCCCTGGAGGAGGATTTTGAAGCAACATCATGATGATTACGGCGCTAAGA 234
 QY 852 cgacgagcgccagcgctcctctgctgagcgccagcgcccgcggtggcgccgcg 911
 Db 233 TGATTTCTGTGATGACATACCTGTGCTGTGACACAGTCCGCTGTGCGAGCCGCCG 174
 QY 912 cgacatgcccgcgcccgcgtgagcatcccgagatcatgacgagcgccgctgagacca 971
 Db 173 AGATATGCCCCCGCGCTGAGTTTCAATACCGAATATCATGCAACCTGTGCTGACACCA 114
 QY 972 cgtgaaactcgtgatacattccgaacctgagcagcagcagcagcagcagcagcagc 1031
 Db 113 TGTAAATATTTGTCAATGATATCCGTAACCTGATGATGAACAGGGGCAATGTCGCG 54
 QY 1032 cctgctgagagcagcgca 1049
 Db 53 CCGTGTGGAAGATGGCGCA 36

RESULT 5
AM211972/c 338 bp mRNA linear EST 03-DEC-1999
LOCUS uc079f01.x1 NCI CGAP Mam3 Mus musculus cDNA clone IMAGE:2648761.3
DEFINITION Similar to SW:RECR_BPPI P06956 RECOMBINASE CRE. [1] ;, mRNA
sequence.
ACCESSION AM211972 GI:6517961
VERSION uc079f01.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 338)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTS: uc079f01.y1
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbdp/image/image.html
MGI:1029213
Seq primer: -400P from Gdbco
High quality sequence stop: 151.
Location/Qualifiers
FEATURES
source
1..338
/organism="Mus musculus"
/strain="129 - C57/B6 - FVBN"
/db_xref="taxon:10090"
/clone="IMAGE:2648761"
/clone_1ib="NCI CGAP Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
BASE COUNT 78 a 93 c 82 g 85 t
ORIGIN
Query Match 14.3%; Score 150.6; DB 9; Length 338;
Best Local Similarity 76.7%; Pred No. 6.3e-11;
Matches 197; Conservative 0; Mismatches 59; Indels 1; Gaps 1;
QY 793 accgggacctgagggcattctcgagggccaccacgcgtatctagcgccgaagasc 852
DB 292 ACTGCGCCCTGAGAGGAT-TTCTAAGCCACATCATCATGATTACGCGCCTAAGAT 234
QY 853 gaacagcgccagcgctactgctgtagagcgccacagcgccgctgtagcgccgcgc 912
DB 233 GACTTTGGTCAGAGATACCTGGCTGTCTGGACACAGTCCGCTGTGAGCGGCTTA 174
QY 913 gacatggcccgccgagcgtgagcattcccgagatcatgagcgccgctggaacac 972
DB 173 CATATGCGCCGCTGAGATTTCATATCCGAGATCATGAACTGTGGCTGACCAAT 114
QY 973 gtgaacactgtgatactacatccgaacttgacagcgagacggcgccatggtgcgc 1032
DB 113 GTAAATATTTGTCAATGAACTATATCCGTAACGTGATAGTGAACAGGGGCAATGTCGC 54
QY 1033 ctgctgtagagagcgcgca 1049

Db 53 CTGCTGGAAGATGGCGCA 37
RESULT 6
AM210918 343 bp mRNA linear EST 03-DEC-1999
LOCUS uc089c09.y1 NCI CGAP Mam3 Mus musculus cDNA clone IMAGE:2649712.5
DEFINITION Similar to SW:RECR_BPPI P06956 RECOMBINASE CRE. [1] ;, mRNA
sequence.
ACCESSION AM210918 GI:6516858
VERSION uc089c09.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 343)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTS: uc089c09.x1
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbdp/image/image.html
MGI:1030164
Seq primer: -40RP from Gdbco
High quality sequence stop: 222.
Location/Qualifiers
FEATURES
source
1..343
/organism="Mus musculus"
/strain="129 - C57/B6 - FVBN"
/db_xref="taxon:10090"
/clone="IMAGE:2649712"
/clone_1ib="NCI CGAP Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
BASE COUNT 95 a 77 c 85 g 84 t 2 others
ORIGIN
Query Match 12.2%; Score 128.2; DB 9; Length 343;
Best Local Similarity 66.5%; Pred No. 5.1e-08;
Matches 228; Conservative 0; Mismatches 110; Indels 5; Gaps 3;
QY 319 cgcgcgaagcgccgctgccccgcacagcgacagcgctgagcctgtagtcgcccgc 378
DB 2 CGTGGGTCCGGGCTGCGCCAGCAAGTGCACGAACTCTGTTACATGCTTATGCGCGG 61
QY 379 atccgcaagagaaacgtgtagcgccgagcgcccaagcgccgctgtagcgccgc 438
DB 62 ATCCAAAGAAAGAAAGTTGATGCCGGTGAACGCTCAAAAGGCTTAGCCTTGAACGC 121
QY 439 accgagcttcgacgaagctgtagcgccgctgtagagcaagcgacgcgctgcgaagatccgc 498
DB 122 ACTGATTTTC-ACCAAGTTCGTTACATGGAATAATACGATTCGTCGACAGATATACGT 180
QY 499 aacctgaccttctctgagca-tgcctacaacaacctctgctgcacatgcgcagatgcgcgc 557

OM of: US-09-662-128a-2 to: GenEmbl.* out_format: pfs

Date: May 10, 2002 4:09 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+2, model -DEV=xlp  
-O=/gen2_1/USPRO.spool/US09662128/runat.08052002.135153.28227/app-query.fasta.1.413  
-DB=genembi -OFMT=fastap -SUFFIX=rgc -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPEXT=4.000  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blonsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09662128.ecgcl.1.7246  
-ICPU=6 -ICPU=3 -LONGIOS -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-662-128a-2  
Query length: 350  
Database: GenEmbl.*  
Database sequences: 1797656  
Database length: 187333701  
Search time (sec): 1991.870000
```

Score list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_pat:AX191674	+ 1797.00	2514.55	1.1e-131	4847	AX191674 Sequence 40 from Pat
gb_pat:AX191653	+ 1797.00	2514.32	1.1e-131	4960	AX191653 Sequence 29 from Pat
gb_pat:AX114854	+ 1797.00	2513.54	1.2e-131	5365	AX114854 Sequence 14 from Pat
gb_sy:AY056050	+ 1793.00	2523.81	3.2e-132	1080	AY056050 Synthetic construct
gb_pat:AX205077	+ 1792.00	2506.68	2.9e-131	5261	AX205077 Sequence 7 from Pat
gb_sy:AF34827	+ 1785.00	2495.67	1.2e-130	5894	AF34827 Cloning vector pTUDC
gb_sy:AF169416	+ 1776.00	2481.77	7.1e-130	6652	AF169416 Cloning vector pACN c
gb_pat:AX150382	+ 1765.00	2484.35	5.1e-130	1074	AX150382 Sequence 1 from Pat
gb_pat:AX191655	+ 1765.00	2484.35	5.1e-130	1074	AX191655 Sequence 1 from Pat
gb_pat:AX191652	+ 1765.00	2470.15	3.1e-129	4491	AX191652 Sequence 23 from Pat
gb_pat:AX191650	+ 1761.00	2478.25	1.1e-129	1125	AX191650 Sequence 18 from Pat
gb_pat:AX191647	+ 1759.00	2469.70	3.3e-129	2004	AX191647 Sequence 13 from Pat
gb_pat:AX150386	+ 1759.00	2469.70	3.3e-129	2004	AX150386 Sequence 5 from Pat
gb_pat:AX191639	+ 1759.00	2469.70	3.3e-129	2004	AX191639 Sequence 5 from Pat
gb_pat:AX191651	+ 1759.00	2469.70	3.3e-129	2004	AX191651 Sequence 20 from Pat
gb_pat:AX343218	+ 1757.00	2473.46	2.1e-129	1032	AX343218 Sequence 17 from Pat
gb_pat:AX348053	+ 1757.00	2473.46	2.1e-129	1032	AX348053 Sequence 1 from Pat
gb_pat:AX353866	+ 1757.00	2473.46	2.1e-129	1032	AX353866 Sequence 32 from Pat
gb_pat:AX353920	+ 1757.00	2469.41	3.5e-129	1553	AX353920 Sequence 1 from Pat
gb_ph:WFLP1CRE	+ 1757.00	2469.41	3.5e-129	1553	WFLP1CRE Bacteriophage phi cre
gb_pat:AX067794	+ 1757.00	2468.28	4.0e-129	1740	AX067794 Sequence 10 from Pat
gb_pat:AX64972	+ 1757.00	2466.88	4.8e-129	2004	AX64972 Sequence 2 from Pat
gb_sy:AF397196	+ 1754.00	2455.84	2.0e-128	6094	AF397196 Retrofiling vector
gb_sy:AF298789	+ 1751.00	2446.18	6.8e-128	6293	AF298789 Recombinase expressid
gb_sy:AF298782	+ 1751.00	2446.02	6.9e-128	6979	AF298782 Recombinase expressid
gb_sy:AF298785	+ 1751.00	2445.92	6.9e-128	6979	AF298785 Recombinase expressid
gb_sy:AF306636	+ 1693.00	2357.21	6.2e-123	14103	AF306636 Plant DNA excision
gb_pat:AX343222	+ 1670.00	2345.61	2.7e-122	1725	AX343222 Sequence 5 from Pat
gb_pat:AX353870	+ 1670.00	2345.61	2.7e-122	1725	AX353870 Sequence 7 from Pat
gb_pat:AX343226	+ 1670.00	2345.61	2.7e-122	1725	AX343226 Sequence 5 from Pat
gb_pat:AX353874	+ 1670.00	2345.19	2.9e-122	1800	AX353874 Sequence 9 from Pat
gb_pat:AX191649	+ 1132.00	1574.22	2.5e-79	5953	AX191649 Sequence 15 from Pat
gb_pat:AR091543	+ 671.00	945.03	2.8e-44	699	AR091543 Sequence 11 from Pat
gb_pat:AR102236	+ 671.00	945.03	2.8e-44	699	AR102236 Sequence 11 from Pat

```
gb_da:AF250878 + 275.00 331.16 4.4e-10 180461 AF250878 Salmonella typhi  
gb_da:STPpHOM1 - 275.00 329.28 5.6e-10 218160 AL513383 Salmonella enter  
gb_da:AE007307 + 206.50 263.04 2.7e-06 10192 AE007307 Shirochizobium meli  
gb_da:PSN5041 - 202.50 253.62 9.1e-06 14907 X98999 Pseudomonas sp. tra  
gb_da:U66917 + 192.00 219.07 0.0008 108845 U66917 Pseudomonas sp. AD
```

seq_name: gb_pat:AX191674

seq_documentation_block:

LOCUS AX191674 4847 bp DNA linear PAT 15-AUG-2001

DEFINITION Sequence 40 from Patent WO0149832.

ACCESSION AX191674

VERSION AX191674.1 GI:15209855

KEYWORDS synthetic construct.

SOURCE synthetic construct.

ORGANISM artificial sequence.

REFERENCE 1 (bases 1 to 4847)

AUTHORS Schwenk, F.

TITLE Transduction of recombinases for inducible gene targeting

JOURNAL Patent: WO 0149832-A 40 12-JUL-2001;

FEATURES location/Qualifiers

source 1..4847

BASE COUNT 1138 a 1238 c 1284 g 1184 t 3 others

ORIGIN /db_xref="taxon:32630"

alignment_scores: Quality: 1797.00 Length: 350

Percent Similarity: 5.134 Gaps: 0

Percent Identity: 100.000

alignment_block: US-09-662-128A-2 x AX191674

Align seg 1/1 to: AX191674 from: 1 to: 4847

1 MetProlylsLysArgLysValSerAspLeuThrValHisGlnAs 17

2775 ATGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2824

17 nleuprolaLeuProValAspAlaThrSerAspGluValArgLysAsn 34

2825 TTTCGTCGATTCACCGGTCGATCGACAGAGATGAGGTTCCAGAACAC 2874

34 eumetaspMetPheArgAspArgGlnAlaPheSerGluHisThrProly 50

2875 TGATGACATCTTACGAGATCGCCAGCGCTTCTTGAGCATACCTGAGAA 2924

51 MetLeuSerValLysArgSerTrpAlaAlaTrpCysLysLeuAsnAs 67

2925 ATGCTTGTGTCGCTTCCCGGTCGTCGCGGCGAGTGGAGTGAATGA 2974

67 naryLysTrpPheProAlaGluProGluAspValAlaGlyTrpLeuLeu 84

2975 CCGGAAATGCTTCCCGCAGAACCTGAAAGTGTTCGATTTATCTTCT 3024

84 yrieGlnAlaArgGlyLeuAlaValLysThrIleGlnIleHisLeuGly 100

3025 ATCTTCAGCGCGCGCGCTGTCGAGTAAACATTCACGACATTTGGCG 3074

101 GlnLeuAsnMetLeuHisArgGlyLeuProArgProSerAspSe 117

3075 CAGCTAACATGCTTCATGCTGCTCCGCGGCGCCAGCCAAAGTACAG 3124

117 rasnAlaValSerLeuValMetArgTrgIleArgLysGluAsnValAsp 134

3125 CAATGCGTTTACCTGCTTATGCGGCGGATCCGAAAGAAACGTTGATG 3174

```

134 IAGLYGIUARGALALYSGLNALALEUALAPHEGLIUAQTHTASPPHEASP 150
|||||
3175 CCGGTGAACGTGCAAAACAGCGCTAGCTGCAACGACACGATTTTGAC 3224
|||||
151 GINVALARGSERLEUMETGLUASERASPARGCYSGINASPILLEARGAS 167
|||||
3225 CAGGTTCTGCTCATGTGAAAAATAGCATGCTGCCAGGATATACGTAA 3274
|||||
167 nleuAlaPheLeuGlyIleAlaTyraThrLeuLeuArgIleAlaGluI 184
|||||
3275 TCTGGCATTTCTGGGATTGCTTATAACACCTGTACGTATACCGAAA 3324
|||||
184 leAlaArgIleArgValLysAspIleSerArgThrAspGlyLysArgMet 200
|||||
3325 TTGCCAGGTTCAGGGTTAAAGTATCTACGACTGACGGGGGAGAAATG 3374
|||||
201 leuIleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyValG 217
|||||
3375 TTTATCCATATTTGGCAGAACGAAAGCGTGGTTACACCGAGGTGTAGA 3424
|||||
217 uLysAlaLeuSerLeuGlyValThrLysLeuValGluArgTrpIleSerV 234
|||||
3425 GAAGGCACTTAGCTGGGGGTAACTAACTGGTCGAGCGATGGATTTTCG 3474
|||||
234 aLserGlyValAlaAspAspProAsnAsnTyrlLeuPheCysArgValArg 250
|||||
3475 TCTCTGTGTAGCTGATGATCCGAAATACCTACTGTTTGGCGGTGAGA 3524
|||||
251 LysAsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAl 267
|||||
3525 AAAAAGGAGTGTGGCCGCGCATCTGGCACACGACGATACCACTCGGC 3574
|||||
267 aLeuGluGlyIlePheGluAlaThrHisArgLeuIleTyrglyAlaLysA 284
|||||
3575 CCTGGAGAGGATTTTGAAGCAACTCATGATTGATTACGGCGCTAAGG 3624
|||||
284 sPASPserGlyGlnArgTyrlLeuAlaTrpSerGlyHisSerAlaArgVal 300
|||||
3625 ATGACTGTGTGACAGATACCTGCTGCTGGACACAGTCAGTCCCGTGTG 3674
|||||
301 GLYALALARGASPMEtAlaArgAlaGlyValSerIleProGluIleLys 317
|||||
3675 GGAGCCGCGGAGATATGGCCGCGCTGGAGTTTCAATACCGAGATCAT 3724
|||||
317 tGlnAlaGlyIleTyrlThrAsnValAsnIleValMetAsnTyrlLeuArg 334
|||||
3725 GCAAGCTGGTGGGTGGACCAATGTAATATTTGCATGAACTAATATCCGA 3774
|||||
334 snLeuAspSerGluThrGlyAlaMetValArgLeuLeuGlnAspGlyAsp 350
|||||
3775 ACCTGGATAGTGAACAGGGGCAATGTGTGCTGCTGGAAAGATGGCGAT 3824
|||||
seq_name: gb_pat:AX191663
seq_documentation_block:
LOCUS AX191663 4960 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 29 from Patent W00149832.
ACCESSION AX191663
VERSION AX191663.1 GI:15209844
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 4960)
AUTHORS Schenk, F.
TITLE Transduction of recombinases for inducible gene targeting
JOURNAL Patent: WO 0149832-A 29 12-JUL-2001;
ARTEMIS Pharmaceuticals GmbH (DE)
FEATURES
source
1..4960
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="vector pCMV-I-Cre-pA"

```

```

BASE COUNT 1225 a 1213 c 1296 g 1226 t
ORIGIN
alignment_scores:
Quality: 1797.00 Length: 350
Ratio: 5.134 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-662-128A-2 x AX191663
Align seg 1/1 to: AX191663 from: 1 to: 4960
1 MetProLysLysArgLysValSerAsnLeuLeuThrValHisGlns 17
|||||
977 ATGCCCAAGAACAGAGAGAGTGTCCAAATTACTGACCGTACCCAAAA 1026
|||||
17 nLeuProAlaLeuProValAlaSerAlaThrSerAspGluValArgLysAsnL 34
|||||
1027 TTTGGCTGCATTACCGGTGCGATGCAACGATGATGAGGTTCGCAAGAAC 1076
|||||
34 eUeLAspPhePheArgAspArgGlnAlaPheSerGluHisThrTrpLys 50
|||||
1077 TGAATGACATGTTCCAGGATCCCGAGCGGTTTCTGACATACCTGGAAA 1126
|||||
51 MetLeuLeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAs 67
|||||
1127 ATGCTTCTGTCCGTTTGGCGGTCTGGGCGCATGTCAGAGTTGAATTA 1176
|||||
67 nArgLysTrpPheProAlaGluProGluAspValArgAspTyrlLeuLeu 84
|||||
1177 CCGAATATGTTTCCGCGAAGCTGAAGATGTTGGCATATATCTTCAT 1226
|||||
84 yrlLeuGlnAlaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuGly 100
|||||
1227 ATCTTCAGGCGCGCGGTCTGCGAGTAAACATNTCCACCACTTTGGGC 1276
|||||
101 GlnLeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSe 117
|||||
1277 CAGCTAAACATGCTTCATGCTGCGTCCGGCTCCACACCAAGTACAG 1326
|||||
117 rAsnAlaValSerLeuValMetArgArgIleArgLysGluAsnValAspA 134
|||||
1327 CAAATGCTGTTTCACTGTATATGGCGGAGATCCGAAACGAAACGTTGAT 1376
|||||
134 IAGLYGIUARGALALYSGLNALALEUALAPHEGLIUAQTHTASPPHEASP 150
|||||
1377 CCGGTGAACGTGCAAAACAGGCTTAGCGTGAACGCACTGATTTGCAC 1426
|||||
151 GlnValArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAs 167
|||||
1427 CAGGTTCTGCTCATGCAAAATACGATCGTGGCAGAGATATACGTAA 1476
|||||
167 nLeuAlaPheLeuGlyIleAlaTyraThrLeuLeuArgIleAlaGluI 184
|||||
1477 TCTGGCATTTCTGGGATTGCTTATAACACCCGTACTACTAAGCGGAAA 1526
|||||
184 leAlaArgIleArgValLysAspIleSerArgThrAspGlyLysArgMet 200
|||||
1527 TTGCCAGATCAGGGTTAAAGATATCTCACGACTACACCGTGGGAGAA 1576
|||||
201 leuIleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyValG 217
|||||
1577 TTTATCCATATTTGGCAGAACGAAAGCGTGGTTACACCGAGGTGTAGA 1626
|||||
217 uLysAlaLeuSerLeuGlyValThrLysLeuValGluArgTrpIleSerV 234
|||||
1627 GAAGGCACTTAGCTGGGGGTAACTAACTGGTCGAGCGATGGATTTTCG 1676
|||||
234 aLserGlyValAlaAspAspProAsnAsnTyrlLeuPheCysArgValArg 250
|||||
1677 TCTCTGTGTAGCTGATGATCCGAAATACCTACTGTTTGGCGGTGACAA 1726
|||||

```

```
251 LysAsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAl 267
|||||
1727 AAAAATGGTGTCCGCCCATCTGCCACAGCAGCAATACATCACTCCGCGC 1776
267 aleuGluGlyIlePheGluAlaThrHisArgLeuIleTyrGlyAlaLysA 284
|||||
1777 CCGTGAAGGATTTTGAAGCAACATCATGATTGATTTACGGCGCTAAGG 1826
284 spAspSerGlyGlnArgTyrLeuAlaIleTrpSerGlyHisSerAlaArgVal 300
|||||
1827 ATACCTCTGTGTCAGAGATACCTGGCTGTGGACACAGTCCGCTGTC 1876
301 GlyAlaAlaArgAspMetAlaArgAlaGlyValSerIleProGluIleMe 317
|||||
1877 GGAGCCGCGCAGATATGCGCCGCTGAGTTTCAATACCGGAGATCAT 1926
317 tGlnAlaGlyGlyTyrPThrAsnValAsnIleValMetAsnTyrIleArgA 334
|||||
1927 GCAGCTGGTGGCTGGACCAATGTAAATATTGTCACTAATATATCCGTA 1976
334 snLeuAspSerGlyThrGlyAlaMetValArgLeuLeuGluAspGlyAsp 350
|||||
1977 ACCTGGATGTGAACAGGGCATGTGCGCTGCTGGAAAGATGGCGAT 2026
seq_name: gb_pat:AX114854
seq_documentation_block:
LOCUS AX114854 5365 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 14 from Patent WO0129208.
ACCESSION AX114854
VERSION AX114854.1 GI:14031796
KEYWORDS
SOURCE
ORGANISM synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 5365)
AUTHORS Kuehn, R., von Melchner, H. and Altschmid, J.
TITLE Conditional gene trapping construct for the disruption of genes
JOURNAL Patent: WO 0129208-A 14 26-Apr-2001;
ARTEMIS Pharmaceuticals GmbH (DE); Frankgen Biotechnologie AG (DE)
FEATURES
Location/Qualifiers
1..5365
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="vector PBADp-pgkCre"
BASE COUNT 1208 a 1448 c 1402 g 1307 t
ORIGIN
alignment_scores:
Quality: 1797.00 Length: 350
Ratio: 5.134 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-662-128A-2 x AX114854 ..
Align seg 1/1 to: AX114854 from: 1 to: 5365
1 MetProLysIleLysArgLysValSerAsnLeuThrValHisGlnAs 17
|||||
1972 ATGCCCAAGAGAAAGAGAGAGTGCATTTAAGACCGTACACCAAA 2021
17 nLeuProAlaLeuProValAspAlaThrSerAspGluValArgLysAsn 34
|||||
2022 TTGGCTGATTACCGGTGATGCCAAGAGTATGAGGTGCAAGAAC 2071
34 euMetAspMetPheArgAspArgGlnAlaPheSerGluHisThrTyrLys 50
|||||
2072 TGATGGACATGTTCAGGGATCCGACGCGCTTTTTCAGCATACCTGGAA 2121
51 MetLeuSerValCysArgSerTrrPAlaIleTrrPcysLysLeuAsnAs 67
|||||
2122 ATGCTTCTGTCCCTTGGCCGGTGTGGCGCATGTCGMAATTGAATA 2171
67 nArgLysTrrPheProAlaGluProGluAspValArgAspTyrLeuLeu 84
|||||
2172 CCGAAATGGTTTCCCGAGAACCTGAAGATGTGCGCATTTATCTTAT 2221
84 yrLeuGlnAlaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuGly 100
|||||
2222 ATCTTCAGCGCGCGGTGTGGCAGCTAATAACTATCCACCAATTTGGC 2271
101 GlnLeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSe 117
|||||
2272 CAGCTAATACATCTTCATCGTGTGCTCCGGGCTCCACAGCAAGTGAC 2321
117 rAsnAlaValSerLeuValMetArgArgIleArgLysGluAsnValAspA 134
|||||
2322 CAATGCTGTTTCAGTGTATATGCGCGGATCCGAAACAAACGTTGATG 2371
134 IagLysGluArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheAsp 150
|||||
2372 CCGGTGAACGTGCMAAACAGGCTCAGCGTTCCAGACACATGATTTGAC 2421
151 GlnValArgSerLeuMetCylasSerAspArgCysGlnAspIleArgAs 167
|||||
2422 CAGGTTCTTCATCTCATGTGAATAATAGCGATCGTCCAGGATATACGTA 2471
167 nLeuAlaPheLeuGlyIleAlaTyrAsnThrLeuLeuArgIleAlaGlu 184
|||||
2472 TCTGCAATTTCTGGGATGTGTTTAAACACCGCTGTACGTATAGCCGAA 2521
184 leuAlaArgIleArgValLysAspIleSerArgThrAspGlyArgMet 200
|||||
2522 TTGCAGAGATCAGGGTTAAAGATATCTCAGCTACTGACGCTGGAGAA 2571
201 LeuIleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyAlaG 217
|||||
2572 TTATATCATATTGGCAGAACGAAACGCTGTGTAGACCGCAGCTGTAA 2621
217 uLysAlaLeuSerLeuGlyValThrLysLeuValGluArgTrrPLeuSer 234
|||||
2622 GAAAGCATTAGCCCTGGGGGTAACTAATGTGTGAGCGCATGATTTCCG 2671
234 alSerGlyValAlaAspAspProAsnAsnTyrIleuPheCysArgValArg 250
|||||
2672 TCTGTGTGTAGCTGATGATGCCAATPACTACCTGTTTGGCCGGTCAGA 2721
251 LysAsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAl 267
|||||
2722 AAAAATGGTGTCCGCCCATCTGCCACACAGCACTACACTCCGCGC 2771
267 aleuGluGlyIlePheGluAlaThrHisArgLeuIleTyrGlyAlaLysA 284
|||||
2772 CCGTGAAGGATTTTGAAGCAACATCATGATTGATTTACGGCGCTAAGG 2821
284 spAspSerGlyGlnArgTyrLeuAlaIleTrpSerGlyHisSerAlaArgVal 300
|||||
2822 ATGACCTGTGTCAGAGATACCTGGCTGTGGACACAGTCCCGCTGTC 2871
301 GlyAlaAlaArgAspMetAlaArgAlaGlyValSerIleProGluIleMe 317
|||||
2872 GGAGCCGCGCAGATATGCGCCGCTGAGATTCAATACCGGAGATCAT 2921
317 tGlnAlaGlyGlyTyrPThrAsnValAsnIleValMetAsnTyrIleArgA 334
|||||
2922 GCAGCTGGTGGCTGGACCAATGTAAATATTGTCTAATGACTATATCCGTA 2971
334 snLeuAspSerGlyThrGlyAlaMetValArgLeuLeuGluAspGlyAsp 350
|||||
2972 ACCTGGATGTGAACAGGGCAATGTGTGCGCTGCTGGAAGATGGCGAT 3021
seq_name: gb_sy:AV056050
```

```

seq_documentation_block:
LOCUS      AY056050                1080 bp    DNA     linear     SYN 19-NOV-2001
DEFINITION Synthetic construct codon-optimized cre recombinase (1cre) gene,
complete cds.
ACCESSION  AY056050
VERSION    AY056050.1 GI:17016299
KEYWORDS
SOURCE     synthetic construct.
           synthetic construct.
           artificial sequence.
REFERENCE  1 (bases 1 to 1080)
AUTHORS   Shimshek,D.R., Kim,J., Huebner,M.R., Spergel,D.J., Buchholz,F.,
           Stewart,F.A., Seeburg,P.H. and Sprengel,R.
TITLE      Codon-Optimized Cre Recombinase Expression in the Mouse
JOURNAL    unpublished
REFERENCE  2 (bases 1 to 1080)
AUTHORS   Shimshek,D.R., Kim,J., Huebner,M.R., Spergel,D.J., Buchholz,F.,
           Stewart,F.A., Seeburg,P.H. and Sprengel,R.
TITLE      Direct Submision
JOURNAL    Submitted (13-SEP-2001) Molecular Neurobiology,
           Max-Planck-Institute, Jahnstr. 29, Heidelberg 69120, Germany
FEATURES
Source     1..1080
           /organism="synthetic construct"
           /db_xref="taxon:32630"
           /note="derived from bacteriophage P1"
misc_feature 1..13
           /note="contains Kozak sequence"
           /gene="13"
gene         14..1069
           /gene="1cre"
           /gene="1cfe"
           /note="contains part of the large T nuclear localization
           signal sequence"
           /codon_start=1
           /transl_table=11
           /product="codon-optimized cre recombinase"
           /protein_id="AAJ31698.1"
           /db_xref="GI:17016300"
           /translation="MWPKKRKRVSNLLTVHONLPALPDATSDVRRKNIMDFRROA
           FSEHTMKMLISVCRSNAAMCKLNNRWKFAPEPDVADLYLLYQAKGLAVKTIQHLGO
           LNMJHRSGLPSPSDSNVAVSLVARRIRKENVAGERAKGLAFERFDQVSLNENS
           DRQDIDNLAFLFIAVNTLRIAEIARIKVDISRTDGGRLIHIGRTLVSTGVE
           KALSLGTYKIEVERNISSGVADDPNNVLCFRVKNVAPASATISOLSTALDIEIEAT
           HRLITGAKDSDGQRIYLAWSHSAHVGAARDMAAGVSIPEIQAGGWTNVTNMYIR
           NLDSEGANVRLLEDGP"
BASE COUNT 248 a 294 c 319 g 219 t
ORIGIN
alignment_scores:
Quality: 1793.00      Length: 350
Ratio: 5.123          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.714
alignment_block:
US-09-662-128a-2 x AY056050
Align seg 1/1 to: AY056050 from: 1 to: 1080
1 MetProLysLysLysArgLysValSerAsnLeuLeuThrValHisGlnAs 17
:::|||||
17 GTGCCCAAGAAAGAGAAAGTCCCAACTGCTGACTGTGCAACCAAAA 66
17 nLeuPrAlaLeuProValAspAlaThrSerAspGluValArgLysAsnL 34
|||||
67 CCTCCCTGCGCTCCCTGTGATGCACCTCTGTGATGAGTCAGAAACACC 116
|||||
34 eumetAspmetPheArgAspArgGlnAlaPheSerGlnHisThrTrpLys 50
|||||
117 TGATGAGCATGTTCAAGGACAGGAGCCCTTCTCTGAACACACCTGGAAG 166
|||||
51 MetLeuLeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuSns 67

```

```

|||||
167 ATGCTCTGCTGTCTGTGCAGATCCTCGGCTGCTGCTGCAAGCTGAACAA 216
|||||
67 nArgLysTrpPheProAlaGluProGluAspValArgAspTrpLeuLeuT 84
|||||
217 CAGGAATGTTCCCTGCTGACCACTGAGATGATGTAGGGAGTAACTCTCTGT 266
|||||
84 yrlLeuGlnAlaArgGlyLeuAlaValLysThrLleGlnGlnHisLeuGly 100
|||||
267 ACCGTCAAGCCAGAGGCTGCTGTGTGAAGCAATCAGAAAGAGAAATGTGATG 416
|||||
317 CAGTCAACATGCTGCACAGAGATCTGGCTGCTGCTGCTGCTGCTGCTGCTG 366
|||||
101 GlnLeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSe 117
|||||
317 CAGTCAACATGCTGCACAGAGATCTGGCTGCTGCTGCTGCTGCTGCTGCTG 366
|||||
117 TAsnAlaValSerLeuValMetArgArgLleArgLysGluAsnValAspA 134
|||||
367 CAATGCTGTGTCTGCTGTGATGAGGAGAAATCAGAAAGAGAAATGTGATG 416
|||||
134 lAgLysGluArgAlaLysGlnAlaLeuAlaPheGluArgTrpAspPheAsp 150
|||||
417 CTGGGAGAGAGAGCCAGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 466
|||||
151 GlnValArgSerLeuMetGluAsnSerAspArgCysGlnAspLleArgAs 167
|||||
467 CAAGTCAGATCCTGATGAGAACTGACAGATGACAGAGACATCCAGAGAA 516
|||||
167 nLeuAlaPheLeuGlyLleAlaTrpAsnThrLeuLeuArgLleAlaGln 184
|||||
517 CTTGGCTTCTCTGCGCATGTGCTTCAACACCTGCTGCTGCTGCTGCTGCTG 566
|||||
184 lLeAlaArgLleArgValLysAspLleSerArgTrpAspGlyGlyArgMet 200
|||||
567 TTGCCAGATCAGAGTGAAGACATCTCCGACCGATGATGGAGAGATG 616
|||||
201 LeuLleHisLleGlyArgTrpThrLysThrLeuValSerThrAlaGlyVal 217
|||||
617 CTGATCCACATTTGCGAGAGCAAGACCTGCTGTCTCCACAGTGGTGTGGA 666
|||||
217 uLysAlaLeuSerLeuGlyValThrLysLeuValGluArgTrpLleSerV 234
|||||
667 GAAGGCGCTGCTCCCGGGGGTTACCAACCTGCTGAGAGAGATGATCTCTG 716
|||||
234 aLseGlyValAlaAspAspProAsnAsnTrpLeuPheCysArgValArg 250
|||||
717 TGTCTGTGTGTGCTGATGACCCCAACAACTACCTGCTGCTGCTGCTGCTG 766
|||||
251 LysAsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAl 267
|||||
767 AAGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 816
|||||
267 aLeuGlnGlyLlePheGlnAlaThrHisArgLeuLleTrpGlyAlaLysA 284
|||||
817 CCTGGAAGGAGATCTTTGAGGCCACCCACCGCTGATGTCATGTGTCGAAAG 866
|||||
284 sPAspSerGlyGlnArgTrpLeuAlaTrpSerGlyHisSerAlaArgVal 300
|||||
867 ATGACTCTGGGACAGACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 916
|||||
301 GlyAlaAlaArgAspMetAlaArgAlaGlyValSerLleProGlnLleMe 317
|||||
917 GGTGCTGCTCCAGGAGACATGGCCAGGGGCTGTGTGTCATCTCCGAAATCAT 966
|||||
317 tGlnAlaGlyGlyTrpThrAsnValAsnLleValMetAsnTrpLleAlaG 334
|||||
967 GCAGGCTGTGTGCTGACCAATGTGAACATAGTGAATGAATCATCATCAGA 1016
|||||
334 snLeuAspSerGluTrnGlyAlaMetValArgLeuLeuGlnLysAsp 350
|||||
1017 ACCTGAGACTGTGAAGACTGGGGCCATGTGATGAGCTGCTGAGATGGGAGC 1066
|||||
seq_name: gb_pat:AX205077

```

```

seq_documentation_block:
LOCUS      AX205077              5261 bp      DNA      linear      PAT 30-AUG-2001
DEFINITION Sequence 7 from Patent WO0155361.
ACCESSION  AX205077
VERSION     AX205077.1  GI:15394318
KEYWORDS
SOURCE      unidentified.
            unclassified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 5261)
AUTHORS     Hardy, S.F.
TITLE       Recombinant aav packaging systems
JOURNAL     Patent: WO 0155361-A 7 02-AUG-2001;
            Chiron Corporation (US)
FEATURES
source      1..5261
            /organism="unidentified"
            /db_xref="taxon:3264"
            /note="recombinant DNA"
BASE COUNT  1309 a 1238 c 1329 g 1385 t
ORIGIN
alignment_scores:
  Quality: 1792.00      Length: 349
  Ratio: 5.135          Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-662-128A-2 x AX205077
Align seg 1/1 to: AX205077 from: 1 to: 5261

2  ProLysLysLysArgLysValSerAsnLeuThrValHisGlnAsnLeu 18
|||||
1201 CCAAGAGAGAGAGAGAGAGTTTCGAATTACTGACCGTACACCAAAATTT 1250
|||||
18  uProLaleuProValAspPalaThrSerAspGluValArgLysAsnLeu 35
|||||
1251 GCCTGCATATCCGGTCGATGCACACAGATGAGGTTCGCAAGAACCTCGA 1300
|||||
35  eLaspMetPheArgAspArgGlnAlaPheSerGluHisThrTrpLysMet 51
|||||
1301 TGGACATGTCAGGGATCGCCAGCGGTTTCTGAGCATACCTCGAAATG 1350
|||||
52  LeuLeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArg 68
|||||
1351 CTTCTGTCGGTTTCCGGTCTGGCGGCATGTCGCAAGTTGAATTAACCG 1400
|||||
68  gLysTrpPheProAlaGluProGluAspValArgAspTrpLeuLeuTrp 85
|||||
1401 GAAATGGTTCCCGCAGAACCTGAAGATGTTCCGCATTAATCTTCTATATC 1450
|||||
85  euGlnAlaArgGlyLeuAlaValLysThrIleGlnHisLeuGlyGln 101
|||||
1451 TTCAGGGCGGGGCTGCGAGTAAACATACTCCAGCAATTTGGCCAG 1500
|||||
102  LeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAs 118
|||||
1501 CTAACATGCTTCATCTCGTCCGGCTCCGACGACCAAGTAGACACAA 1550
|||||
118  nAlaValSerLeuValMetArgArgIleArgLysGluAsnValAspAlaG 135
|||||
1551 TGCCTGTTCACTGGTTATGCGGGGATCCGAAAAAGAAAGAGTTGATCCG 1600
|||||
135  LysGluArgAlaLysGlnAlaLeuAlaPheGluArgTrpAspPheAspGln 151
|||||
1601 GTGACGCTGCAAAACAGGCTCTAGCGTTGCAACGACATGATTCGACAG 1650
|||||
152  ValArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAsnLe 168
|||||
1651 GTTCGTTCACTCATGAAATAGCATCGCTGCCAGATATACGTAATCT 1700
|||||

seq_name: gb_sy.AF334827
seq_documentation_block:
LOCUS      AF334827              5894 bp      DNA      circular SYN 20-FEB-2001
DEFINITION Cloning vector pTurbo-Cre, complete sequence.
ACCESSION  AF334827
VERSION     AF334827.1  GI:12965137
KEYWORDS
SOURCE      Cloning vector pTurbo-Cre.
ORGANISM    Cloning vector pTurbo-Cre
            artificial sequence; vectors.
REFERENCE   1 (bases 1 to 5894)
AUTHORS     Lu, Z.H., Graubert, T.A. and Ley, T.J.
TITLE       Direct Submission
JOURNAL     Submitted (10-JUN-2001) Division of Oncology, Section of Stem Cell
            Biology, Washington University School of Medicine, 660 South Euclid
            Avenue, Campus Box 8007, St. Louis, MO 63110-1093, USA
FEATURES
source      1..5894
            /organism="Cloning vector pTurbo-Cre"
            /db_xref="taxon:152298"
            1..364
            /note="cytomegalovirus immediate early region"
            365..646
            /note="chicken beta-actin"
            647..1619
            /note="chicken beta-actin"
            1620..1664
            /note="rabbit beta-globin"

```

|||||.....
2098 TCGCTTTCACCTGTTGTGCGCGGATCCGAAAGAAAGCTGATGCCG 2147
135 IYGLIARGLAIALYSGIINALAENALAPHEGILUARGTHRASPHASPGIN 151
2148 GTGAACGTGCAAAACAGGCTCTAGCGTTCCGACGACACTATTTCACAG 2197
152 VALARGSERLEUMETGIUANSERASPARCYSGIINASPILARGASULE 168
2198 GTTCTTACTCATGAGAAATAGCGATCGCTCCAGAGATATACGTACT 2247
168 VALAPHELEUGLYILEALATYRASNTHLEULEUARGIIIEALGIULEA 185
2248 GGCATTTCGGGGATTCCTTATACCCCTGTTACGATAGCGGAATTG 2297
185 IARGTILEARGVALLYASPILESERARGTHRASPGIYGLIARGMETLEU 201
2298 CCAGAGATCAGGGTTAAAGATATCTACGTAAGTGAAGGTGGAGATGTTA 2347
202 ILEHISILEGLIARGTHRLYSTHLEUVALSERTHRALGIYALIGULY 218
2348 ATCCATTATTCGCGAGAACGAAACGCTGGTTAGACCCGACAGTGTAGAGA 2397
218 SALALEUSERLEUGLYVALTHRLYSLEUVALGIUARGTTPILLESVALS 235
2398 GGCACCTTACCCTGGGGGTACTAAACCTGTGACCGATGGAATTTCCGTCT 2447
235 ERGLIYALIALASPARSPROASNAANTYRLEUPHECYASARGVALARGLYS 251
2448 CTGTGTACTCATGATGATCGAATACACTACCTGTTTTCGCGGGTCAGAAAA 2497
252 ASNGIYVALIALAIALAPROSERALATHSERGILNLEUSERTHIRARGALALE 268
2498 AATGGHTTGGCGCGCATCTCCACCGCCAGCAGCTATCACTACGTGCGCCT 2547
268 UGLIUGLYIIEPHEGLIUALATHRHISARGLEUIETRYGIALALYSASPA 285
2548 GGAGGGATTTTGAAGCAACTCACTACGATGATTATTACCGCCGTAAAGATG 2597
285 SPARGIYGLINARGTYRLEUALATPSPERGIIYHISERALIAARGVALGILY 301
2598 ACCTGTGTCAGAGATACCTGGCCTGTGTGGACACAGTGCCTGCTGGGA 2647
302 ALAIALARGASPMETHIARGIAGLYALSERTILEPQIUILEMETGI 318
2648 GCGCGCGGAGATWGGCCGCGCTGAGTTTCAATACCGAGATACATGCA 2697
318 NALAGIYGLYTRPTHRASNVALASNILEVALMETASNTYRILEARGANL 335
2698 AGCTGGTGGCTGGACCAATGTAATTTGTCTATGAACTATATCCGTAAAC 2747
335 EUANSPERGIIUTHRGYIAMIETVALARGLEULEUGIIASGIIASPGI 350
2748 TGGATTAGTGAACAGGGGGCAATGGTGGCTCTCTCTCTCTCTCTCTCTCT 2794
seq_name: gb_sy:AF169416

seq.document.location: block: 6652 bp DNA linear SYN 04-AUG-1999

LOCUS AF169416
DEFINITION Cloning vector pACN complete sequence.
ACCESSION AF169416
VERSION AF169416.1 GI:5690439
KEYWORDS
SOURCE .
ORGANISM Cloning vector pACN.
Cloning vector pACN
Artificial sequence: vectors.
REFERENCE 1 (bases 1 to 6652)
AUTHORS Bunting,M., Bernstein,K.E., Greer,J.M., Capechi,M.R. and Thomas,K.R.
TITLE Targeting genes for self-excision in the germ line
JOURNAL genes Dev. 13 (12), 1524-1528 (1999)
MEDLINE 99315626
REFERENCE 2 (bases 1 to 6652)

AUTHORS Bunting,M., Bernstein,K.E., Greer,J.M., Capecchi,M.R. and Thomas,K.R.
 TITLE Direct Submission
 JOURNAL Submitted (15-JUL-1999) Human Genetics, University of Utah, 15 North 2030 East Room 5440, SLC, UT 84112-5331, USA
 FEATURES Location/Qualifiers
 Source 1.6652

/organism="Cloning vector pACN"
 /db_xref="taxon:100905"
 complement(682..715)
 /note="loxP element"
 polyA_signal complement(742..789)
 /note="HSV thymidine kinase polyA signal"
 gene complement(797..1600)
 /gene="neo"
 complement(797..1600)
 /gene="neo"
 CDS complement(797..1600)
 /gene="neo"
 /codon_start=1
 /product="neomycin phosphotransferase"
 /protein_id="AAD47089.1"
 /db_xref="GI:5690441"
 /translation="MGAIEDDGLHAGSPAAMVERLFGYDAAQOTICSDPAVRLSA
 OGRPVLFVKTDLSGALNEDDEARLSLMTATGVPFCAVLDVTEAGRWLLLEVP
 ODLSHSLPAKVSITMADAMRLTLPATCPEDHOKHRIEARTRMELGVDOD
 LDEHOGIAPALFRLKAKMPDGEDLVVTHGDCLPIMVENGFRSGFTDCGRIGVA
 DRYODIALATRDIAELGGEWADRLVLYGIAAPSDSRIAEYRLDEFF"

promoter complement(1617..2323)
 /note="RNA polymerase II large subunit promoter"
 complement(2339..2386)
 /note="HSV thymidine kinase polyA signal"
 gene complement(2533..3651)
 /gene="cre"
 complement(101n(2533..2736,2803..3651))
 /gene="cre"
 /codon_start=1
 /product="Cre recombinase with nuclear localization
 sequence"

CDS complement(2737..2802)
 /gene="cre"
 /note="SV40 t-antigen intron"
 complement(3652..4350)
 /note="testis specific angiotensin-converting enzyme
 promoter (tACE)"
 promoter (tACE)
 complement(4358..4391)
 /note="loxP element"

misc_feature complement(1781 c 1700 g 1578 t
 ORIGIN
 intron complement(2737..2802)
 /gene="cre"
 /note="SV40 t-antigen intron"
 complement(3652..4350)
 /note="testis specific angiotensin-converting enzyme
 promoter (tACE)"
 promoter (tACE)
 complement(4358..4391)
 /note="loxP element"

misc_feature complement(4358..4391)
 /note="loxP element"
 BASE COUNT 1593 a 1781 c 1700 g 1578 t
 ORIGIN

alignment_scores:
 Quality: 1776.00 Length: 372
 Ratio: 5.074 Gaps: 1
 Percent Similarity: 94.086 Percent Identity: 94.086

alignment_block:
 US-09-662-128a-2 x AF169416/rev ..

Align seg 1/1 to reverse of: AF169416 from: 1 to: 6652

1 MetProLysLysLysArgLysValSerasnLeuThrValHisGlnAs 17
 3651 ATGCCCAAGAGAGAGAGAGGTGTCATTTACTGACCGACCAACAA 3602
 17 nLeuProAlaLeuProValAspAlaThrSerAspGluValArgLysAsn 34

|||||
 3601 TTGGCTGCATTACCGGTGATGCAACAGATGATGAGTTGCAAGAAC 3552
 34 eUMetAspMetPheArgAspArgGlnAlaPheSerGluHisThrLys 50
 3551 TATGAGACATGTTCCAGGATCGCCAGGCGTTTCTCGACATACCTGGA 3502
 51 MetLeuSerValLysArgSerThrAlaAlaTrpCysLysLeuAsnAs 67
 3501 ATGCTTCGTGCGTTGGCGGTCCGTGGCGGCGCATGTGCAGATTGAAT 3452
 67 nArgLysTrpPheProAlaGluProGluAspValArgAspTyrLeuLeu 84
 3451 CCGGAATGGTTTCCCGCAACCTGAAGATGTTCCGATATCTCTAT 3402
 84 YLeuGlnAlaArgGlyLeuAlaValLysTrpTrpLeuGlnHisLeuGly 100
 3401 ATCTTCAGCGCGCGGTGTCGACATGAATACTATCCAGCAACTTGGGC 3352
 101 GlnLeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAsp 117
 3351 CAGCTAAACATGCTTCATCGTCCGCGTCCGCGTCCGACAGACAGTACAG 3302
 117 rAsnAlaValSerLeuValMetArgGlyLeuGlyGluAsnValAsp 134
 3301 CAATGCTGTTCACTGGTTATGCGCGGATCCGAAAGAAAGCTTGATG 3252
 134 IagIyGluArgAlaLysGlnAlaLeuAlaPheGluArgTrpAspPheAsp 150
 3251 CCGGTGAACGTCGAAACAGGCTCTACGTTCCGACAGCAGCTGTTCCGAC 3202
 151 GlnValArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAs 167
 3201 CAGGTGCTTCATCATGAGAAATAGAGATCGCTCCAGATATACGTA 3152
 167 nLeuAlaPheLeuGlyIleAlaTyrAsnThrLeuLeuArgIleAlaGlu 184
 3151 TCTGCACTTCTCGGGATTCCTATTAAACACCGCTTACGATATGCCGAA 3102
 184 LeuAlaArgIleArgValLysAspIleSerArgThrAspGlyArgMet 200
 3101 TTGCGAAGATCAGGTTAAAGATATCTCAGTACGACGCGTGGAGAAAG 3052
 201 LeuIleHisIleGlyArgTrpThrLysThrLeuValSerThrAlaGlyVal 217
 3051 TTATATCATATTGGCAGAACGAAACGCTGTGTAGACCGCAGGTGAGA 3002
 217 uLysAlaLeuSerLeuGlyValThrLysLeuValGluArgTrpIleSer 234
 3001 GAAGCAGCTTACCGCTGGGGTAACTAACTGTCGACGAGATGATTTCCG 2952
 234 alSerLysValAlaAspAspProAsnAsnTyrLeuPheCysArgValArg 250
 2951 TCTCTGTGTAGCTGATGATCCGAATTACTACTGTTTCCGGGTGAGA 2902
 251 LysAsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAl 267
 2901 AAAAATGCTGTCGCGCGCATCTCCACACCGCATATCACTACCTCCGCG 2852
 267 alLeuGluGlyIlePheGluAlaThrHisArgLeuIleTyrGlyAlaLys 283
 2851 CCTGGAAGGATTTTGAAGCAACTCATCATGATTGATTACGCGGCTAAG 2802
 283 283
 2801 TAAATATAAATTTTAAAGTATAAGTGTAAACTACTGATCTTAAT 2752
 284 AspAspSerGlyGlnArgTyrLeuAlaTrpSerG 295
 2751 GTTGTGATTTTAGAGATGACTGTGTCAGAGATACCTGCGCTGCTGG 2702
 295 yHisSerAlaArgValGlyAlaAlaArgAspMetAlaArgAlaGlyValAs 312
 |||||

```

2701 ACACAGTGGCCGTGTGCGAGCCGCGGAGATATGCCCCGCTGAGATT 2652
312 erlleprogluilemetglnalaglytrphtasnvalasnileval 328
|||||
2651 CAAATCCGAGATCATCAAGCTGTGGTGGCGCAATTAATATTTGTC 2602
329 Metasntylleatrgasnleuaspsergluthrghylametvalargle 345
|||||
2601 ATGAACATATTCGCTACCTGATAGTGAACAGGGCAATGTTGGCCT 2552
345 ulleugluaspolyasp 350
|||||
2551 GCTGGAAGATGGCGAT 2536

seq_name: gb_pat:AX150382

seq_documentation block:
LOCUS AX150382 1074 bp DNA linear PAT 25-JUN-2001
DEFINITION Sequence 1 from Patent EP118668.
ACCESSION AX150382
VERSION AX150382.1 GI:14571608
KEYWORDS
SOURCE
.
ORGANISM
.
synthetic construct.
artificial sequence.
1 (bases 1 to 1074)
REFERENCE
Schwenk,F.
AUTHORS
TITLE
Patent: EP 118668-A 1 25-JUL-2001;
JOURNAL
ARTEMIS Pharmaceuticals GmbH (DE)
FEATURES
location/Qualifiers
source
1..1074
/oranism="synthetic construct"
/db_xref="taxon:32630"
/note="DNA sequence coding for a fusion protein Tat-Cre"
1..1074
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC42784.1"
/db_xref="GI:14571609"
/translation="MGYGRKKRRQRRRNSNLLIVHQNLPALPDATSDVEKRLMDM
FRDROAFSEHTMRLSLVCRSMAMWCKLNNRKFPAEPEDVDYLLVYQAKLAVKI
OOHLGOI.NMLHRRSGLRPSDSNAVSLVMRRIKENVADGERAKALAEKRTDQVR
SLMENSROCDIRNLAFILGLIAYNLILAEIARIKVDLSRTDGRMLIHRTDTLV
STAGVEKALSLIGVTKLVERMISVSGVADDPNNVLFCRVKRNGLVAPASATISOLSTALE
GIEPATRIITYGAKKDSGORYLXMSGHSARVGAARDMANAGVSIDEIMQAGMTVNI
VMNTRMLDSEGTAMRRLLEDGP"
BASE COUNT 268 a 254 c 299 g 253 t
ORIGIN

alignment_scores:
Quality: 1765.00 Length: 348
Ratio: 5.086 Gaps: 0
Percent Similarity: 99.713 Percent Identity: 98.563

alignment block:
US-09-662-128a-2 x AX150382 ..

Align seq 1/1 to: AX150382 from: 1 to: 1074

3 LysLysLysArgLysValSerAsnLeuLeuThrValHisGlnAsnLeuP 19
|||||
28 CAACGCCGCGCGGATGTCATTTACTGACCGGTACACCAAAATTTGGC 77
19 oAlaLeuProValAspAlaThrSerAspGluValArgLysAsnLeuMet 36
|||||
78 TGCATTTCCGGTCGATGCAACGAGATGAGGTTCGCAAGAACCTGATG 127
36 spMetPheArgAspArgGlnAlaPheSerGluHisThrTrpLysMetLeu 52
|||||
128 ACATGTTCCAGGATCGCCAGCGCTTTTCTGAGCATACGTGAAAATGCTT 177

```

```

53 LeuSerValCysArgSerTrpAlaIleTrpCysLysLeuAsnArgLys 69
|||||
178 CTGTCCGTTTCCGCGTGTGCGGCGATGTCAGTTCATTAACCGGAA 227
69 sTrpPheProAlaGluProGluAspValArgAspTrpLeuLeuTrpLeu 86
|||||
228 ATGTTTCCCGCAGAACCTGGAAGATGTTCCGATATATCTTATATCTTC 277
86 lnaIaArgLysLeuAlaValLysThrIleGlnGlnHisLeuLysLeu 102
|||||
278 AGGCGCGCGTCTGCGCATTAACCTATCCAGCAACATTGGGCCAGCTA 327
103 AsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAsn 119
|||||
328 AACATGCTTATCTGCTCGCTCGGGCTGCCACGACCAAGTACGCAATGC 377
119 aValSerLeuValMetArgArgIleArgLysGluAsnValAspAlaGly 136
|||||
378 TGTTCACCTGTATTCGCGGAGATCCGAAAGAAAGGTTGATGCGGATG 427
136 lnaArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheAspGlnVal 152
|||||
428 AACGTCAAAACAGGCTCTACGCTTCAACGCACTGATTTCCAGCAGGTT 477
153 ArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAsnLeuAl 169
|||||
478 CGTTCACCTCATGGAATAATACGATCGCTGCCAGAGATACGTATCTGCG 527
169 aPheLeuGlyIleAlaLysAsnThrLeuLeuArgIleAlaGluIleAla 186
|||||
528 ATTCTGTGGGATTCCTTATACACCCCTGTTACTGATACGCAATATGCCA 577
186 rGlyIleArgValLysAspIleSerArgThrAspGlyIleArgMetLeu 202
|||||
578 GGATCAGGTTAAAGATATCTCACGTACTGACGCTGAGGAAGATTATATC 627
203 HisIleGlyArgThrLysThrLeuValSerThrAlaGlyAlaGluLysAl 219
|||||
628 CATATTGGCAGACGAAACGCTGTTAGCCACGCAAGTGTAGAGAGGC 677
219 aLeuSerLeuGlyValThrLysLeuValGluArgTrpIleSerValSerG 236
|||||
678 ACTTACCTCGGGGTAACATAACTGCTGACGATGATTTCCGCTCTG 727
236 lValAlaAspAspProAsnAsnTrpLeuPheCysArgValArgLysAsn 252
|||||
728 GTGTAGCTGATGATCCGATTAACCTACTGTTTCCGGGTCAGAAAAAT 777
253 GlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAlaLeuG 269
|||||
778 GGTGTTGCCGCGCCATCTGCCACGACGCAAGCATCACTCGCGCTGGA 827
269 uGlyIlePheGluAlaThrHisArgLeuIleTyrGlyValLysAspAsp 286
|||||
828 AGGATTTTGAAGCAACTCATGATGATTATTCGCGCTAAGGATGACT 877
286 eArgLysLysArgTrpLeuAlaThrSerGlyHisSerAlaArgValGlyAla 302
|||||
878 CTGGTCAGAGATACCTGGCTCTGCTGACACAGTCCCGGTGGAGGCC 927
303 AlaArgAspMetAlaArgAlaGlyAlaSerIleProGluIleMetGlnAl 319
|||||
928 GCCCGAGATATGCCCCGCTGAGTTCAATACCGGAGATCATGCAACG 977
319 aGlyIleTrpThrAsnValAsnIleValMetAspTrpIleArgAsnLeu 336
|||||
978 TGGTGGCTGACCAATGTAATATGTCATGAACTATATCCCTAACCTGG 1027
336 spSerGluThrGlyAlaMetValArgLeuLeuGluLysPoiLysp 350
|||||
1028 ATATGTAAACAGGCGCAATGCTGCGCTGCTGGAAGATGGCGAT 1071

```


Integrated dna sequence
Patent: WO 0179512-A 23 25-OCT-2001;

Iconix Pharmaceuticals Inc. (US)

FEATURES

Location/Qualifiers
1..4491
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="vector"

BASE COUNT 1160 a 1065 c 1119 g 1147 t
ORIGIN

alignment_scores:

Quality: 1765.00 Length: 348
Ratio: 5.086 Gaps: 0
Percent Similarity: 99.713 Percent Identity: 98.563

alignment_block:

US-09-662-128a-2 x AX347590 ..

Align seg 1/1 to: AX347590 from: 1 to: 4491

```

3 LysLysLysArgLysValSerAsnLeuThrValHisGlnAsnLeuP 19
:::.....::: |||||.....
169 GGTCAAGCCGCTGCGCTCCAAATTACTGACGCTACACCAAAATTGGC 218
19 cAlaLeuProValAspAlaThrSerAspGluValArgLysAsnLeuMet 36
|||||.....
219 TGCATTACCGGTCGATGCACAGAGTGAAGAGTTCCGCAAGAACCTGATGG 268
36 sPmetPheArgAspArgGlnAlaPheSerGluHisThrTrpLysMetLeu 52
|||||.....
269 ACATGTTCCAGGATCCAGCGCGTTTCTGACATCACTGGAAAAATGCTT 318
53 LeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArgLys 69
|||||.....
319 CTGTCCGTTTGGCGGTGCTGCGGCGCATGTCGAAGTTGAATTAACCGGAA 368
69 strPheProAlaGluProGluAspValArgAspTrpLeuLeuTrpLeuG 86
|||||.....
369 ATGGTTTCCCGCAGACCTGAAAGATGTTCCGATTAATCTTCTATATCTTC 418
86 lnaAlaArgLysLeuAlaValLysTrpHisGlnHisLeuGlnLeu 102
|||||.....
419 AGCGCGCGGCTCGCAGTAAACATATCCACACATTTGGGCCAGCTA 468
103 AsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAsnAl 119
|||||.....
469 AACATGCTTCATCGTCGTCGCGGCTCCAGCAAGAGTGCAGCAATATGC 518
119 aValSerLeuValMetLarGArgLLeuArgLysGlnAsnValAspAlaGly 136
|||||.....
519 TGTTCCTCTGTTATGCGGGGATCCGAAAAAGAAACGTGATGCCGGTG 568
136 lnaArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheAspGlnVal 152
|||||.....
569 AACGTCAAAACAGCGCTAGCTGCAAGCAGCATGATTTCCACCAAGTT 618
153 ArgSerLeuMetGluAsnSerAspArgCysGlnAspLLeuArgAsnLeuAl 169
|||||.....
619 CGTTCACCTACATGAAATAGCGATCGCTGCCAGATATACGTAATCTGGC 668
169 aPheLeuGlyLLeuAlaTrpAsnThrLeuLeuArgLLeuAlaGluLeuAla 186
|||||.....
669 ATTTCCTGGGATGCTTATAACACCGCTTACGATATAGCGAAATGGCA 718
186 rGllLeuArgValLysAspLLeuSerArgThrAspGlyGlnArgMetLeuLe 202
|||||.....
719 GGATCAGGGGTTAAAGATATCTCAGCTACGCGTGGGAGAAATGTTATATC 768
203 HisLLeuGlyArgThrLysThrLeuValSerThrAlaGlyValGluLysAl 219
|||||.....
769 CATATTGGCAGACGAAACCGCTGTTAGACCCGCAAGCTGTACGAAAGGC 818

```

seq_name: gb_pat:AX191652

seq_documentation_block:

LOCUS AX191652

1125 bp DNA linear PAT 15-AUG-2001

DEFINITION Sequence 18 from Patent WO0149832.

ACCESSION AX191652

1125 bp DNA linear PAT 15-AUG-2001

VERSION AX191652.1

1125 bp DNA linear PAT 15-AUG-2001

KEYWORDS

1125 bp DNA linear PAT 15-AUG-2001

ORGANISM

1125 bp DNA linear PAT 15-AUG-2001

REFERENCE

1 (bases 1 to 1125)

AUTHORS

Schwenk, F.

TITLE

Transduction of recombinases for inducible gene targeting

JOURNAL

Patent: WO 0149832-A 18 12-JUL-2001;

ARTEMIS

Pharmaceuticals GmbH (DE)

FEATURES

Location/Qualifiers

SOURCE

1..1125

CDS

1..1125

ORIGIN

278 a 265 c 316 g 266 t

alignment_scores:

278 a 265 c 316 g 266 t

Quality: 1761.00 Length: 348
 Ratio: 5.075 Gaps: 0
 Percent Similarity: 99.713 Percent Identity: 98.276
 alignment block:
 US-09-662-128a-2 x AX191652 ..
 Align seg 1/1 to: AX191652 from: 1 to: 1125

```

3 LysLysLysArgLysValSerAsnLeuThrValHisGlnAsnLeuPr 19
  :::::::::::|::|::|::|::|::|::|::|::|::|::|::|::|
28 CAACGCCGCCGCCGCGCATGTCACATTACTGACCGTACACCAAAATTGCC 77
  |||
19 oAlaLeuProValAspAlaThrSerAspGluValArgLysAsnLeuMet 36
  |||
78 TGCATTACCGGTCGATGCACACAGTGAAGTTCGCAAGAACCTGATGG 127
  |||
36 sPmPheArgAspArgGlnAlaPheSerGlnHisThrTrpLysMetLeu 52
  |||
128 ACATGTTCCAGGATCCGACGCGGTTTCTGAGCATACCTGGAAAATGCTT 177
  |||
53 LeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArgLy 69
  |||
178 CTGTCCGTTTGGCGGTCTGCGGCGGCATGTCAGATTGAATTAACCGGAA 227
  |||
69 sTrpPheProAlaGluProGluAspValArgAspTyrLeuLeuTyrLeug 86
  |||
228 ATGGTTTCCCGACAACTGAAGATGTCGCGATTATCTTATATCTTC 277
  |||
86 LAlaArgGlyLeuAlaValLysThrIleGlnHisLeuGlnLeu 102
  |||
278 AGCGCGCGGCTCTGACACTAAAACCTTCCAGCAACATTGGGCGCACCTA 327
  |||
103 AsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAsnAl 119
  |||
328 AACATGCTTCATCGTCGCGTCCGCGGCTCCACGACCAAGTACAGCAATGC 377
  |||
119 aValSerLeuValMetArgIleArgLysGlnAsnValAspAlaGly 136
  |||
378 TGTTCACCTGTTATGCGGCGGATCCGAAAAGAAAGATTGATGCCGCTG 427
  |||
136 LuAlaGlyLysGlnAlaLeuAlaPheGluArgThrAspPheAspGlnVal 152
  |||
428 AACGTGCAAAACAGCTCTAGCGTTGCAACGACATGATTTCGACCGAGTT 477
  |||
153 ArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAsnLeuAl 169
  |||
478 CGTTCACCTCATGAAAATAGCATCGCTGCGAGATATACGTAATCTGCC 527
  |||
169 aPheLeuGlyIleAlaTyrAsnThrLeuLeuArgIleAlaGluIleAla 186
  |||
528 ATTTCTGGGATTCCTTATTAACCCCTGTTACGTATACCCGAAATTGGCA 577
  |||
186 rGluLeuArgValLysAspIleSerArgThrAspGlyLysArgMetLeuIle 202
  |||
578 GGAACAGGGTTAAAGATATCTACGTACTGACGCTGGAGAAATGTAATC 627
  |||
203 HisIleGlyArgThrLysThrLeuValSerThrAlaGlyValGluLysAl 219
  |||
628 CATATTGGCAGACGAAACCGCTGTTACACCGCGGTGAGAAAGGC 677
  |||
219 aLeuSerLeuGlyValThrLysLeuValGluArgTrpIleSerValSerG 236
  |||
678 ACTTACCGCGGGGTAACTAACTGCTGACGAGATGATTCGCTCTCTG 727
  |||
236 LyValAlaAspAspProAsnAsnTyrLeuPheCysArgValArgLysAsn 252
  |||
728 GTGTACTGATGATCCGAATTAACCTGTTTTCGCGGCTCAGAAAATAT 777
  |||
253 GlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAlaLeuG 269
  |||
778 GGTGTTCGCGCGCATCTGCCACAGCCAGCATATCACTCGCGCCCTTGA 827
  |||

```

```

269 uGlyIlePheGluAlaThrHisArgLeuIleTyrGlyAlaLysAspAsp 286
  |||
828 AGGATTTTGAAGCAACTCATCGATGATGATTACGGCCCTAAGATGACT 877
  |||
286 eGlyGlnArgTyrLeuAlaTrpSerGlyHisSerAlaArgValGlyAla 302
  |||
878 CTGTGAGAGATACCTGCGCTGCTGGACACAGTGCCTGTCGGAGCC 927
  |||
303 AlaArgAspMetAlaArgAlaGlyValSerIleProGluIleMetGlnAl 319
  |||
928 GCGCGAGATATAGCCCGCGCTGACTTTCATATCCGAGATCATGCAACG 977
  |||
319 aGlyGlyTyrThrAsnValAsnIleValMetAsnTyrIleArgAsnLeu 336
  |||
978 TGTGCTGCTGACCAATGTAATATATGTCATGACATATATCCGTAACTCG 1027
  |||
336 sPSeGlnThrGlyValMetValArgLeuLeuGluAspGlyAsp 350
  |||
1028 ATAGTGAACAGGGGCAATGGTGGCGCTGTGGAAGATGGCGAT 1071
  |||

```

```

seq_name: gb_pat:AX191650
seq_documentation_block:
LOCUS AX191650 4727 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 16 from Patent WO0149832.
ACCESSION AX191650
VERSION AX191650.1 GI:15209838
KEYWORDS
SOURCE
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 4727)
AUTHORS Schwenk,F.
TITLE Transduction of recombinases for inducible gene targeting
JOURNAL Patent: WO 0149832-A 16 12-Jul-2001;
ARTEMIS Pharmaceuticals GmbH (DE)
FEATURES
source 1, 4727
location/Qualifiers
BASE COUNT 1172 a 1232 c 1177 g 1146 t
ORIGIN

```

```

alignment_scores:
Quality: 1761.00 Length: 348
Ratio: 5.075 Gaps: 0
Percent Similarity: 99.713 Percent Identity: 98.276
alignment block:
US-09-662-128a-2 x AX191650/rev ..

```

```

Align seg 1/1 to reverse of: AX191650 from: 1 to: 4727

3 LysLysLysArgLysValSerAsnLeuThrValHisGlnAsnLeuPr 19
  :::::::::::|::|::|::|::|::|::|::|::|::|::|::|::|
1274 CAACGCCGCCGCCGCGCATGTCACATTACTGACCGTACACCAAAATTGCC 1225
  |||
19 oAlaLeuProValAspAlaThrSerAspGluValArgLysAsnLeuMet 36
  |||
1224 TGCATTACCGGTCGATGCACACAGTGAAGTTCGCAAGAACCTGATGG 1175
  |||
36 sPmPheArgAspArgGlnAlaPheSerGlnHisThrTrpLysMetLeu 52
  |||
1174 ACATGTTCAAGGATGCCGCGGCTTTCGACATACCTGGAAAATGCTT 1125
  |||
53 LeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArgLy 69
  |||
1124 CTGTCCGTTTGGCGGTCTGCGGCGCATGTCAGATTGAATTAACCGGAA 1075
  |||
69 sTrpPheProAlaGluProGluAspValArgAspTyrLeuLeuTyrLeug 86
  |||

```

```

1074 ATGCTTCCGCGAAGACCTGAGATGTCGCGATTACTTCTATATCTTC 1025
      1025
86 1n1a1aargglyleuvalaValIystr11leGlnGlnHisLeuGlyGlnLeu 102
      1025
1024 AGGCGCGCGGCTCTACAGTAAAAACTATCCAGACATTTGGCCAGCTA 975
      975
103 AsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAsn1 119
      119
974 AACATGCTTATCGTCGTCGCGCTCCACGACCAAGTACGCAATGC 925
      925
119 aValSerLeuValMetArgArgIleArgLysGlnAsnValAspAlaGly 136
      136
924 TGTTCACCTGTTATGCGCGGATCCGAAAAAGAAAGCTTGATCGCGGTG 875
      875
136 1uArgAlaLysGlnAlaLeuAlaPheGlyuArgThrAspPheAspGlnVal 152
      152
874 AACGTCGCAAAACAGGCTCTACGCTTCAGACGCACTGATTTCCGACAGTT 825
      825
153 ArgSerLeuMetGlnAsnSerAspArgCysGlnAspIleArgAsnLeuAl 169
      169
824 CGTTCACCTGAGAAATAGCGAGATCGCTGCCAGATATACGTATATCTGGC 775
      775
169 aPheLeuGlyIleAlaIyAsnThrLeuLeuArgIleAlaGlnIleAla 186
      186
774 ATTCTGGGGAATGCTTATACACCCCTGTATAGCTATAGCCGAATTCGA 725
      725
186 rglIeArgValIyAspIleSerArgThrAspGlyIyArgMetLeuIle 202
      202
724 GGATGAGGTTAAAGATATCTCAGCTACTGAGGTGGAGATATTTATC 675
      675
203 HisIleGlyArgThrIystrIleuValSerThrAlaGlyValGluLysAl 219
      219
674 CATATTTGCGAGACGAAACCCGCTGTAGCACCCGAGGTGTAGAGAAAGGC 625
      625
219 aLeuSerLeuGlyValIyThrLysLeuValGluArgTrpIleSerValSerG 236
      236
624 ACTATGCGCTGGGTAACCTGCTGACGAGATGATTCGCTCTG 575
      575
236 1yValAlaIAspAspProAsnAsnIyLeuPheCysArgValArgLysAsn 252
      252
574 GTGTACTGATGATCGAATATCTACTGTTTGGCGGGTCAGAAAAAT 525
      525
253 GlyValAlaIAspSerAlaThrSerGlnLeuSerThrArgAlaLeuGly 269
      269
524 GGTGTTGCGCGCATCTGCCACAGCAGCATATCACTCGCGCCCTGGA 475
      475
269 uGlyIlePheGlnAlaThrHisArgLeuIleTrpGlyAlaLysAspAsp 286
      286
474 AGGGATTTTGAAGCAATCATCATGATGATTTCAGCGCTAAGATGACT 425
      425
286 eArgIyGlnArgIyLeuAlaTrpSerGlyHisSerAlaArgValGlyAla 302
      302
424 CTGTGACGAGATATCTGGCTGGTGTGGACACAGTCCGCTGCGAGACC 375
      375
303 AlaArgAspMetAlaArgAlaGlyValSerIleProGluIleMetGlnAl 319
      319
374 GCCCGAGATATGCGCGCTGAGATTTCATACCGGAGATATATCAAGC 325
      325
319 aGlyIyTrpThrAsnValAsnIleValMetAsnIyIleArgAsnLeu 336
      336
324 TGGGGCTGAGCAATGTAAATTTCTCATGAACTATATCCGTAACTG 275
      275
336 sPserGluThrIyAlaMetValArgLeuGlnAspGlyAsp 350
      350
274 ATAGTGAACAGGGGCAATGTGCGCTGCGTGAAGATGGCGAT 231
      231
seq_name: gb_pat:AX191647
seq_documentation_block:
LOCUS AX191647 1623 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 13 from Patent WO0149832.
ACCESSION AX191647
VERSION AX191647.1 GI:15209835

```

```

KEYWORDS
SOURCE      synthetic construct.
ORGANISM    artificial sequence
REFERENCE   1 (bases 1 to 1623)
AUTHORS     Schwenk, F.
TITLE       Transduction of recombinases for inducible gene targeting
JOURNAL     Patent: WO 0149832-A 13 12-JUL-2001;
            Artemis Pharmaceuticals GmbH (DE)
            Location/Qualifiers
FEATURES
    source
        1..1623
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="DNA sequence coding for a fusion protein
            deltaVP22core-StrepTag"
            1..1620
            /note="unnamed protein product"
            /codon_start=1
            /transl_table=11
            /protein_id="CAC51200.1"
            /db_xref="GI:15209835"
            /translation="MASMTGGQMGCRDPSTAPTRSKTPDAQGLARKLHFSAPRNPDA
            KTPRVAGFNKRVFCAVGRILAAHMAHMAAVALMDMSRPREDDELINELLGITTRVYC
            EGNKLLOANELVNPDDVADATATGRSASAPTEPRPARSARRRVEGTE
            LGSIPVMNSADIQHSGRMSNLTIVHON.PALPYDATSDVFKNLMDPRROAFS
            EHTWRKLSVCSMAKRNKLNKRWKPEPEPDVDRDYLILYLOARGLAVKTIQOHLGOLN
            MLHRSGLEPRSDSNVSNVLMRRIRKENVNDGERAKOLAEFRDPDQVRSIMNSDR
            CODIRNLAVLGLAIVNTLRLAEIARIKVD.ISRTDGGMLIHIGRTLVSTAGVEKA
            LSGVTKIVLEWTVSVGVAADPNNTLFCFRVKNQVAASATSSQLSTRALGIFBATIR
            LIYGARDSDGGRYGLAMSGHSAFVGAARMARAGVSIPEIMQAGMTVNTVMNTIRNL
            DSEFGAMVLEIDGDIERGSAVMAHRHPOFG"
BASE COUNT      368 a      440 c      471 g      344 t
ORIGIN
alignment_scores:
    Quality: 1759.00      Length: 344
    Ratio: 5.113      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 99.419
alignment_block:
US-09-662-128A-2 x AX191647 ..
Align seg 1/1 to: AX191647 from: 1 to: 1623
7 LysValSerAsnLeuLeuThrValHisGlnAsnLeuProAlaLeuProva 23
      1025
538 CGCATGTCCTCAATTTACTGACCGCTACACCAAAATTTGCTGCAATACCGGT 587
      587
23 1AspAlaThrSerAspGlyValArgLysAsnLeuMetAspMetPheArg 40
      1025
588 CGATGCAACGAGTGAAGGTTCCGACAGACCTGATGACATGTTCAGAGG 637
      637
40 sPArgGlnAlaPheSerGluHisThrTrpLysMetLeuLeuSerValCys 56
      1025
638 ATCCGACAGCGGTTTCTGAGCATACCTGGAATAATCTTCTGCTGTTGC 687
      687
57 ArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArgLysTrpPheProAl 73
      1025
688 CGGTCTGTGGGCGGATGTCAGATTGAAATTAACCGGAAATGTTTCCCGC 737
      737
73 aGluProGluAspValArgAspIyLeuLeuTrpLeuGlnAlaArgGlyL 90
      1025
738 AGAACCTGAAGATGTCGCGATATCTCTATATCTTCAGCGCGCGGTTC 787
      787
90 euAlaValIystrIleGlnGlnHisLeuGlyGlnLeuAsnMetLeuHis 106
      1025
788 TGGCAGTAAAACTATCCAGCAATTTGGCCAGCTAAACATGCTTCAT 837
      837
107 ArgArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuVa 123
      1025
838 CGTCGCTCGGCGCTCCACGACCAAGTACAGCAATGCTGTTCACTGCT 887
      887

```

```

123 lmetargargilleargylsuaasvalaspaalaglyluarqalalysg 140
|||||
888 TATGGGGGGATCCGAAAGAAAGCTTTGATGCCGGTGAACGTGCAGAAC 937
|||||
140 lnaalaleualaaphegluarqthrarspheaapgluvalarqserleumet 156
|||||
938 AGGCTCTAGCGCTTCGAACGACGATGATTTTCAGCAGGTTGCTTCACTCATG 987
|||||
157 gluanserarsparqcygsluasplleargasnleualaphleuglyl 173
|||||
988 GAAATATGCGATCGCTCCAGGATATACGTAATCTGCAATTTCTGGGGAT 1037
|||||
173 ealatyrasnthrleuualargilleaagluilealaarqileargval 190
|||||
1038 TGCTTATTAACACCTGTAGCTATAGCCGAATATGCCAGATCAGGGTTA 1087
|||||
190 ysasplleserargthrarspglglyargmetleuilehisileglyarq 206
|||||
1088 AAGTATCTACACTACTGACGGTGGGGAATGTATATCATATTGGCAGA 1137
|||||
207 ThrlysthrleuualserthralsaglyvalglulysalaleuSerleugl 223
|||||
1138 ACAGAAACGCTGGTGTAGCACCGAGCTGTAGAGAAAGCCTTACCTGGG 1187
|||||
223 yvalthrlyleuvalgluarqtrpileservalserglyalalaaspa 240
|||||
1188 GGTAACTAAAGTGTGACGAGATGATTTCCCTCTCTGTGTAGCTGATG 1237
|||||
240 sPrCnaasnaantyrleupheCysArgvalArqlyAsnGLyAlaAla 256
|||||
1238 ATCCGAAATACCTACCTGTTTCCGGGTCACAGAAATGCTGTGGCCG 1287
|||||
257 ProSerAlaThrSerGlnleuSerThrargalaleugllylephagl 273
|||||
1288 CCATCTGCACACGACGACGATCACTGCGCCCTGGAGAGGATTTTGA 1337
|||||
273 uAlaThrHlsArgleuiletyrGLyAlaLysAspAspserglylnargt 290
|||||
1338 ACCAACTCATGATGATTACGGGCTCAAGATGACTGTGCTAGAGAT 1387
|||||
290 yfLeuAlaTrpSerGlyHisSerAlaArqvalglYalaAlaArqspmet 306
|||||
1388 ACCTGGCTGGTGTGACACAGTCCCGTGTCCGAGCCGCGGAGATG 1437
|||||
307 AlaArqalaglyValserlleProgluilemetGlnaaglyltyrpth 323
|||||
1438 GCCCGCGCTGAGTTCAATCCGAGATCATGCAAGCTGTGGCTGGAC 1487
|||||
323 rAsnValAsnIleValmetAsnTyrIleArqAsnleuAspSerGlyThrg 340
|||||
1488 CAATGTAAATTTCTCATGAACTATATCCGTAACCTGATAGTGAACAG 1537
|||||
340 lYalameValArqleuGluaspglyasp 350
|||||
1538 GGCAATGAGTGGCGCTCTCGAAGATGGCGAT 1569
|||||
seq_name: gb_pat:AX150386

```

seq_documentation_block:

LOCUS AX150386 2004 bp DNA linear PAT 25-JUN-2001
DEFINITION Sequence 5 from Patent EP1118668.
ACCESSION AX150386
VERSION AX150386.1 GI:14571612
KEYWORDS
SOURCE
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 2004)
AUTHORS Schwenk, F.
TITLE Transduction of recombinases for inducible gene targeting
JOURNAL Patent: EP 1118668-A 5 25-JUN-2001;
ARTEMIS Pharmaceuticals GmbH (DE)
FEATURES
Location/Qualifiers

```

source
1. 2004
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="DNA sequence coding for a fusion protein VP22-Cre"
1. 2004
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC42792.1"
/db_xref="GI:14571613"
/translation="MTRSRYSVSGSPREVPDEYEDLYTPSSGMAIPDPSPTSRGA
LQTRRQGEVRFQYDESDYALYGSSSDDEDEPRPVSAGAVLSGEPAPAP
PPAGSGAGRPPTTAPRPTORVATKAPARAETTRKSAQPEASALPDAPST
APTRSKTPAOGTLARKLHSTRAPRPDPAPTRVAGEKRVCAAVGRALAAARAAV
OLMMSRPRTDEDLNEILGTTTIRVTCBEKLNLANELVNDVVQDDAATATGR
LPALEVDATSDVEVRKLNMDMFRDQAFSEHTMKMLSVCSRMAAMCKLNRRKMPAP
EDVADYLLYLOARCLAVKTIQOHLGOLNMLHRSGLPSPDSNAVSLVMRRIRKED
AGERAKQALAFERTDFDQVRSLSMNSPDODIRNLAFICAYVTLIRIETIRKVD
ISRDGRMLIHIGRTKTLVSTAGVEKALSIGYTKLVERHISVSGVADDPNNYLCRV
RKNGVAPSAATSLRPALEGFIEATHRLITGAKDSDGRTYLAWSGHSARVGAARDMA
RAGVSTPEIMQAGCTVNVATVMTIRNLDSRTGAMVRLLEDG"
BASE COUNT 419 a 608 c 592 g 385 t
ORIGIN

```

alignment_scores:

Quality:	1759.00	Length:	344
Percent Similarity:	100.000	Gaps:	0
Ratio:	5.113	Percent Identity:	99.419

alignment_block:

US-09-662-128a-2 x AX150386 ..

Align seg 1/1 to: AX150386 from: 1 to: 2004

```

7 LysValSerAsnleuThrValHisGlnAsnleuProAlaLeuProVa 23
|||||
970 CGCATGCTCCAAATTTCATGACCGTACACCAAAATTTGCTGCTATACCGGT 1019
|||||
23 lAspaAlaThrSerAspGluValArqlyAsnleuMetAspMetArq 40
|||||
1020 CGATGCAACGAGTATGAGTGTCCAGAAACCTGATGACATGTTCCAGG 1069
|||||
40 sPARqGlnAlaPheserGlnHisThrTrpLysMetleuLeuSerValcys 56
|||||
1070 ATCCGCAAGCGCTTTCTGAGCATACCTGGAATAATGCTTCTGCTCGTTGC 1119
|||||
57 ArgSerTrpAlaAlaTrpCysLysleuAsnArqLysTrpPheProAl 73
|||||
1120 CGGTCTGTGGGGCGCATGTGCAAGTTCAATACCGGAATAGGTTTCCCGC 1169
|||||
73 agLupProGluAspYalArqAspTyrleuLeuTyrleuGlnAlaArqGLY 90
|||||
1170 AGAACCTGAAGATGTTCCGATATATCTTCAATCTTCCAGCGCGGGGTC 1219
|||||
90 euaAlaValLysThrIleGlnGlnHisleuGlyGlnleuAsnmetleuHis 106
|||||
1220 TGGCAGTAAAAACTATCCACACATTTGGCCAGCTAAACATGCTTCAT 1269
|||||
107 ArgArgSerGlyLeuProArgProSerAspSerAsnAlaValSerleuVa 123
|||||
1270 CGTGGTCCGGGCTGCCAGACCAAGTGAACCAATGCTGTTTCACTGGT 1319
|||||
123 lmetargargilleargylsuaasvalaspaalaglyluarqalalysg 140
|||||
1320 TATGGCGCGGATCCGAAAGAAAGCTTGTATGCCGGTGAACGTCGCAAAC 1369
|||||
140 lnaalaleualaaphegluarqthrarspheaapgluvalarqserleumet 156
|||||
1370 AGGCTCTAGCGCTTCGAACGACGATGATTTTCAGACGAGTTCCTACTCATG 1419
|||||
157 gluanserarsparqcygsluasplleargasnleualaphleuglyl 173
|||||

```

```

|||||
1420 GAAATAGGATCGTGGCAGATATACGATATCGGCAATTCGGGAT 1469
173 eaIatYrAsnThrLeuLeuArgIleAlaGluIleAlaArgIleArgValL 190
1470 TGGTTATACACCCCTGTACGTATAGCCGAAATTCACGAGATCAGGGTTA 1519
190 ysAspIleSerArgThrAspGlyArgMetLeuIleHisIleGlyArg 206
1520 AAGATATCTCAGCTACTGACGGTGGAGATGTTATATCATATTTGGCAGA 1569
207 ThrLysThrLeuValSerThrAlaGlyAlaGluLysAlaLeuSerLeuG 223
1570 ACGAAACCGTGGTTAGCACCGCAGGTAGAGAAAGGCACTTAGCCGTGG 1619
223 yAlaThrLysLeuValGluArgTrpIleSerValSerGlyValAlaAspA 240
1620 GGTAACTAAACCTGCTGAGGATGATTCCTGCTGCTGATGCTGATG 1669
240 sPrAsnAsnTrpLysLeuPheCysArgValArgLysAsnGlyAlaAla 256
1670 ATCCGATTAACCTACCTGTTTGGCGGGTCAGAAAAAATGTTGGCCGCG 1719
257 ProSerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheG 273
1720 CCATCTGCCACACCGCAGCTATCACTCGCGCCCTGGAGAGGATTTTGA 1769
273 uAlaThrHisArgLeuIleTyrGlyAlaLysAspAspSerGlyGlnArgT 290
1770 AGCAATCATGATGATGATTACGGCGCTAAGATGACTGTGTCAGAGAT 1819
290 yLeuAlaIleTrpSerGlyHisSerAlaArgValGlyAlaAlaArgAspMet 306
1820 ACCTGGCTGCTGTGGACACAGTCCGCTGCGAGCGCGCAGGATATG 1869
307 AlaArgAlaGlyValSerIleProGluIleMetGlnAlaGlyIleTrpTh 323
1870 GCCCGCGCTGGAGTTTCAATACCGAGATCATGCAAGCTGATGGCTGAC 1919
323 rAsnValAsnIleValMetAsnTrpIleArgAsnLeuAspSerGluThrG 340
1920 CAATGTAATATTTGTCATGAACCTATATCCGTAACCTGATGTAACAG 1969
340 yAlaMetValArgLeuGluAspGlyAsp 350
1970 GGGCAATGCTGCGCTGCTGGAAGATGGCGAT 2001

seq_name: gb_pat:AX191639
seq_documentation_block:
LOCUS AX191639 2004 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 5 from Patent WO0149832.
ACCESSION AX191639
VERSION AX191639.1 GI:15209828
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
1 (bases 1 to 2004)
AUTHORS
Schweik, F.
TITLE
Transduction of recombinases for inducible gene targeting
JOURNAL
Patent: WO 0149832-A 5 12-Jul-2001;
ARTEMIS Pharmaceuticals GmbH (DE)
FEATURES
location/Qualifiers
source
1..2004
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="DNA sequence coding for a fusion protein VP22-Cre"
1..2004
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC51197.1"

```

```

/db_xref="GI:15209829"
/translation="MTRSRVSKSPGREVPDEYEDLYTPSSGNASPDSPDTRRGA
LQTRSNQREVEVVOYDESDVLYGSSSEDEHEVETRRPRPVSGAALSGPGARAP
PPAGSGAGKRTPTTAPRAPRORVATATAPAPAEITRRGRSAQPESAALPDAPAST
APPRSKTPAGOLAKRLHFEETAPRNPDAPEPTPVAGFNKRVCAAGRLAAMAAVAY
QIWDMSRPRTDEDLDELGITITRTVVCCKMLDRAHEIVNPVOVDATATRGR
SAASRPTERRAPARSRRPRREGETEIGSTSPVWMSADIQHSGGMSNLTRYHON
LPALPYDATDEYVRKMLMDMFRDQAFSEHTWKMLLYCSRWAAMCKLNKKRPAPB
EDVRDLYLYQAGLAVKTQQHLQGLNMLHRRSGLPKPSDNAVSLVWRIRKENV
AGERAKOALAEFRDTPDOYRSLMENDRCODIRNLAFGLIAYNLIRLAEIARIRVK
ISRTDGRMLIHGRKTLVSTAGVEKALISYTKLIVERMISVSGVADDPNNYLRCRV
RKNGVAPSATQSULSTRALEGTETAEATHRLITGAKDSGORYLAWSGHSARVQADMA
RAGVSIPEIMQAGWTNVNIVMTIRNLSERIGAMVRLLEDQ"
BASE COUNT 419 a 608 c 592 g 385 t
ORIGIN

```

```

alignment_scores:
  quality: 1759.00      length: 344
  ratio: 5.113          gaps: 0
Percent Similarity: 100.000    Percent Identity: 99.419

```

```

alignment_block:
US-09-662-128A-2 x AX191639

```

```

Align seq 1/1 to: AX191639 from: 1 to: 2004

```

```

7 LysValSerAsnLeuLeuThrValHisGlnAsnLeuProAlaLeuProVa 23
970 CGCATGTCCAATTTACGACCGTACACCAAAATTTCCGCAATTCACCGGT 1019
23 LAspAlaThrSerAspGluValArgLysAsnLeuMetAspMetPheArg 40
1020 CGATGCAACGAGTGATGAGGTGTCGAAAGACCTGATGACATCTTCAGGG 1069
40 sPArgGlnAlaPheSerGlnHisThrTrpLysMetLeuLeuSerValCys 56
1070 ATCCGACAGCGCTTTCTGAGCATACCTGAAATGCTGCTGCTGCTTCC 1119
57 ArgSerTrpAlaAlaIleTrpCysLysLeuAsnAsnArgCysTrpPheProAl 73
1120 CGGTGCTGGCGGCGATGTTGCAATTAACCGGAAATGTTTCCGCC 1169
73 aGluProGluAspValArgAspTrpLeuLeuTrpLeuGlnAlaArgGlyL 90
1170 AGAACCTGAAGATGTTGCGATATCTTCTATATCTTCAGGCCGCGGCTC 1219
90 euAlaValLysThrIleGlnGlnHisLeuGlyGlnLeuAsnMetLeuHis 106
1220 TGGCAGTAAACCTATCCAGACATTTGGCGCAGCTAAACATGCTTCAT 1269
107 ArgArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuVa 123
1270 CGTGGCTCGGGCTGCGACGACCAAGTGCAGCAATGCTGTTCCACGTG 1319
123 lMetArgArgIleArgLysGluAsnValAspAlaGlyGluArgAlaLysG 140
1320 TATGCGCGGATCCGAAAAAGAAACGTTGATCCGGTGAAGTGCAGAAC 1369
140 lAlaIleuAlaPheGluArgTrpAspPheAspGlnValArgSerLeuMet 156
1370 AGGCTTACGCTTCGACACCACTGATTTGACACAGGTCGTTCACTCATG 1419
157 GluAsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyI 173
1420 GAAATATAGCGATGCTGCAGAGATATACGATATTCGGCATTTCTGGGAT 1469
173 eaIatYrAsnThrLeuLeuArgIleAlaGluIleAlaArgIleArgValL 190
1470 TGGTTATACACCCCTGTACGTATAGCCGAAATTCACGAGATCAGGGTTA 1519
190 ysAspIleSerArgThrAspGlyArgMetLeuIleHisIleGlyArg 206
|||||

```

```
1520 AAGATATCTCAGCTACTGACGTGGAGGAAATGTTAATCCATATATGGCAGA 1569
207 ThrIysThrIeuValSerThrAlaGlyValGluLysAlaIeuSerLeuGI 223
|||||
1570 ACGAAAACGGTGGTTAGCACCGGAGGTGTAGAAAGGCACTTAGCCTGGG 1619
223 yValThrLysIeuValGluArgTrpIleSerValSerGIyValAlaAspA 240
|||||
1620 GGTAACTAAACTGTCGACGAGATGATTCGCTCTGTGTAGCTGATG 1669
240 sPProAsnAsnTyrlIeuPheCysArgValArgLysAsnGIyValAlaA 256
|||||
1670 ATCCGAATTAAGTAACTGTTTGGCGGTCAGAAAAAATGTGTGCCGCG 1719
257 ProSerAlaThrSerGlnIeuSerThrArgAlaIeuGluGIyIlePheGI 273
|||||
1720 CCATCTGCCACCGCCAGCTATCAACTCGCCCTGGAAAGGATTTTGA 1769
273 uAlaThrHisArgLeuIleTyrglyAlaLysAspAspSerGIyGlnArgT 290
|||||
1770 AGCACTCATCGATGATTACGGCGCTAAGGATGACTGTGGTCAGAGAT 1819
290 yrlIeuAlaTrpSerGIyHisSerAlaArgValGIyAlaAlaArgAspMet 306
|||||
1820 ACCTGGCTGGTCTGACACACAGTCCCGTCTGGAGCCGCGGAGATATG 1869
307 AlaArgAlaGlyValSerIleProGluIleMetGlnAlaGIyTrpTh 323
|||||
1870 GCCCGGGTGGAGTTCAATACCGAGATCATGCMAAGCTGTGGCTGGAC 1919
323 rAsnValAsnIleValMetAsnTyrlIeArgAsnIeuAspSerGIuThrg 340
|||||
1920 CAATGTAATATTTGTCATGAACTATATCCGTACCTGGATAGTGAACAG 1969
340 lyAlaMetValArgLeuIeuGluAspGIyAsp 350
|||||
1970 GGGCAATGTGTGGCTCTCGAAGATGGCGAT 2001
```


OM of: US-09-662-128a-2 to: N_Geneseq_032802:* out_format : pfs

Date: May 10, 2002 4:15 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frimast_pzn.model -DEV=xlp  
-Q/cgcn2_1/US09662128/runat_08052002.135154.28269/app-query.fasta.1.413  
-DB=N_Geneseq_032802 -OPMT=fastp -SUFFIX=rrg -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELDP=6.000 -DELDP=7.000 -START=1 -MATRIX=blsnum62  
-TRANS=human4.0.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFORMAT=pfs  
-NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09662128 -ECGN1_1.595 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT_THREADS=1
```

Search information block:

```
Query: US-09-662-128a-2  
Query length: 350  
Database: N_Geneseq_032802:*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 257.240000
```

score_list:

```
Sequence Strd Orig ZScore EScore Len ! Documentation  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH41175 + 1797.00 3456.81 3.4e-184 1  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09280 + 1797.00 3438.78 3.0e-183 4  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09269 + 1797.00 3438.52 3.1e-183 4  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH04528 + 1797.00 3438.65 3.5e-183 5  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAH19900 + 1792.00 3438.43 3.6e-183 22  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH06331 + 1792.00 3429.18 1.2e-182 5  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAH09257 + 1765.00 3394.57 9.9e-181 1  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAH16712 + 1765.00 3378.63 7.6e-180 44  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09265 + 1761.00 3386.30 2.9e-180 1  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09263 + 1759.00 3378.34 7.9e-180 1  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09259 + 1759.00 3376.00 1.1e-179 2  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09268 + 1759.00 3375.72 1.1e-179 2  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09266 + 1759.00 3367.02 3.4e-179 4  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH62541 + 1758.00 3380.03 6.4e-180 11  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH10215 + 1757.00 3379.51 6.4e-180 11  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH58065 + 1757.00 3373.69 1.4e-179 17  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAH05702 + 1757.00 3372.12 1.8e-179 20  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH10217 + 1757.00 3370.37 2.2e-179 2  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAH05703 + 1757.00 3359.74 8.6e-179 60  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH81258 + 1756.00 3377.37 9.0e-180 1  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH10216 + 1754.00 3364.36 4.6e-179 2  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH10219 + 1754.00 3364.56 4.6e-179 2  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH62547 + 1741.00 3347.10 4.4e-178 11  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH62545 + 1739.00 3343.22 7.2e-177 11  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH62542 + 1737.00 3339.35 1.2e-177 11  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH62543 + 1729.00 3337.41 1.5e-177 11  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH62546 + 1729.00 3332.85 6.6e-177 11  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAH19899 + 1658.00 3187.83 6.3e-176 11  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH09264 + 1713.00 3149.99 2.3e-171 5  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAH07268 + 671.00 1279.92 6.0e-63 69  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAH21209 + 120.00 125.95 114.89 16  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH53965 + 116.00 201.98 0.0067 8  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH81364 + 106.00 182.61 0.0799 8  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH84553 + 106.00 182.61 0.0799 8  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH52553 + 106.00 182.61 0.0799 8  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH53443 + 105.50 181.75 0.0892 8  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH56133 + 104.00 178.73 0.1314 8
```

```
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH55857 - 104.00 145.13 9.77  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAH33565 + 103.50 176.46 0.1757  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABH13509 + 101.00 155.80 2.49  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABH18790 + 101.00 155.07 2.73  
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH41223 + 100.50 105.51 1.6e+03
```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH41175

seq_documentation_block:

ID AAH41175 standard; DNA: 1050 BP.

AAH41175:

23-AUG-2001 (first entry)

Mammalian Cre recombinase gene.

Mammalian Cre recombinase; organ transplantation; gene therapy;

animal disease model; ds.

Mammalia.

JP2001086989-A.

03-APR-2001.

17-SEP-1999; 99JP-0264364.

17-SEP-1999; 99JP-0264364.

(OSAU) UNIV OSAKA.

WPI: 2001-360321/38.

P-PSDB; AAB98695.

Mammalian type Cre recombinase gene modified so as to have a high
expression efficiency in a mammal by selecting a codon of high
frequency for use in the mammal for expressing Cre recombinase -

Claim 2; Page 7-9; 12pp; Japanese.

The present sequence is a mammalian type Cre recombinase gene. This
sequence can be modified so as to have a high expression efficiency in a
mammal by selecting a codon of high frequency for use in the mammal for
expressing Cre recombinase. The gene can be used in organ
transplantation, gene therapy and creation of animal disease models.

Sequence 1050 BP; 204 A; 383 C; 339 G; 124 T; 0 other;

alignment_scores:

```
Quality: 1797.00 Length: 350  
Ratio: 5.134 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

US-09-662-128a-2 x AAH41175 ..

Align seg 1/1 to: AAH41175 from: 1 to: 1050

```
1 MetProlyLysLysArgLysValSerAsnLeuLeuThrValHisGlnAs 17  
|||||  
1 ATGCCAGAGAGAGAGAGAGAGAGTGCAGCACTGCTGACCGCGACAGAA 50  
|||||  
17 nleuProbiaLeuProValaAspAlaThrSerAspGluValArgLysAsn 34  
|||||  
51 CTTGCCCGCCCTGCGCCGTCGAGCCACCGACGAGAGAGTGGCGAAGAAC 100  
|||||  
34 eumetAspMetPheArgAspArgGlnAlaPheSerGluHisThrTrpLys 50  
|||||  
101 TGATGAGACATGTTCCCGGACCGCGACGCGCTTCACAGCAGCACCTGAG 150  
|||||  
51 MetLeuLeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAs 67
```

```

|||||
151 ATGCTGCTGAGCGTGTGCGGAGCTGGGCGCTGTGTGCAAGCTGAACA 200
67 nArglyStrpPheProAlaGluProGluAspValArgAspTyrLeuLeu 84
|||||
201 CCGCAAGTGTGTTCCCGCGAGCCCGAGAGAGCTGGCGCATCTGCTGT 250
84 yrlLeuGlnAlaArgGlyLeuAlaValLysThrLleGlnGlnHisLeuGly 100
|||||
251 ACCTGCAAGCCCGCGGCTGGCGCGTGAAGACCATCCAGACAGCACTGGGC 300
101 GlnLeuAsnMetLeuHisArgSerGlyLeuProArgProSerAspSe 117
|||||
301 CAGCTAAACATGCTGCACCGCGCGAGCGGCTGCCCGCCGACGAGACAG 350
117 rAsnAlaValSerLeuValMetArgArgLleArgLysGluAsnValAspA 134
|||||
351 CAACGCCGTGAGCTGTGATGCGCGCATCCGCAAGAGAGACGTGAGAG 400
134 laGlyLysLysArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheasp 150
|||||
401 CCGGCGAGCGCGCAAGCAGGCGCTGGCTTGAGGCGCACGAGACTTCGAC 450
151 GlnValArgSerLeuMetGluAsnSerAspArgCysGlnAspLleArgAs 167
|||||
451 CAGGTGCGCAGCGCTGATGGAGAAACAGCAGACCGCTGCCAGACATCCG 500
167 nLeuAlaPheLeuGlyLleAlaTyrAsnThrLeuLeuArgLleAlaGlu 184
|||||
501 CCTGGCTTCCTGGGCATCGCTACCAACACCTGTCGCGCATCGCGAAGA 550
184 leAlaArgLleArgValLysAspLleSerArgThrAspLysGlyArgmet 200
551 TCGCGCGCATCCCGGTGAAGAGACATCAGCCGCGACGCGCGCGCGCAG 600
201 LeuLleHisLleGlyArgThrLysThrLeuValSerThrAlaGlyValG 217
601 CTGATCCACATCGCGCGCACCAAGACCCGTGTGAGCAGCGCGCGCTGGA 650
217 uLysAlaLeuSerLeuGlyValThrLysLeuValGluArgTyrLleSerV 234
651 GAAGGCGCTGAGCTGGCGGTGACCAAGCTGGTGAGCGCTGTGATACGG 700
234 alSerGlyValAlaAspAspProAsnAsnTyrLeuPheCysArgValArg 250
701 TGAGCGCGCGTGGCGAGAGACCCCAACAATCACTGTCTTCGCCGTGGC 750
251 LysAsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAl 267
751 AAGAACGCGCTGGCGCGCCCGCCAGCCAGCCAGCTGAGCACCGCGGC 800
267 alLeuGluGlyLlePheGluAlaThrHisArgLeuLleTyrGlyAlaLys 284
801 CCTGGAGAGCATCTTCGAGGCCACCCACCGCTGATCTACGCGGCCAAG 850
284 spAspSerGlyLysArgTyrLeuAlaTyrSerGlyHisSerAlaArgVal 300
851 ACGACAGCGCGCAGCGCTACCTGCTGAGCGGCCACAGCGCGCGCTG 900
301 GlyAlaAlaArgAspMetAlaArgAlaGlyValSerLleProGluLle 317
901 GGGCGCGCGCGCAGCATGGCGCGCGCGGTGAGCATCCCGCAGANCAT 950
317 tGlnAlaGlyLleTyrThrAsnValAsnLleValMetAsnTyrLleArg 334
951 GCGAGCGCGCGCGCTGGACCAACTGGAACATCGTATGAATCAATCCGCA 1000
334 snLeuAspSerGluThrGlyAlaMetValArgLeuLeuGluAspGlyAsp 350
1001 ACCTGGACAGAGAGACCGCGCATGTGTGGCGCTGTGAGAGAGCGGAGC 1050
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAD09280

```

```

seq_documentation_block:
ID      AAD09280 standard; DNA; 4847 BP.
XX
AC      AAD09280;
XX
DT      12-SEP-2001 (first entry)
XX
DE      pGK-cre-pA vector DNA.
XX
KW      DNA recombinase domain; protein transduction domain; PTD;
KW      gene alteration; fusion protein; Human immunodeficiency virus;
KW      HIV; pGK-cre-pA vector; ds.
XX
OS      Unidentified.
XX
PN      WO200149832-A2.
XX
PD      12-JUL-2001.
XX
PF      05-JAN-2001; 2001WO-EP00060.
XX
PR      07-JAN-2000; 2000EP-0100351.
PR      10-NOV-2000; 2000EP-0124595.
XX
PA      (ARTE-) ARTEMIS PHARM GMBH.
XX
PI      Schwenk F;
XX
DR      WPI; 2001-441873/47.
XX
PT      Using site-specific DNA recombinase domain/protein transduction domain
PT      fusion proteins for inducing target gene alterations in organisms or
PT      cell cultures
XX
PS      Example; Page 67-68; 85pp; English.
XX
CC      The present invention relates to use of fusion proteins comprising
CC      a site-specific DNA recombinase domain e.g. Cre and a protein
CC      transduction domain (PTD) e.g. the Human immunodeficiency virus
CC      (HIV) derived TAT peptide, for preparing an agent for inducing
CC      target gene alterations in a living organism or cell culture. The
CC      present invention also provides a method for inducing gene
CC      alterations in living organisms using the fusion proteins of the
CC      invention. The present sequence is a pGK-cre-pA DNA.
XX
SQ      Sequence 4847 BP; 1139 A; 1238 C; 1283 G; 1184 T; 3 other;

alignment_scores:
      Quality: 1797.00      Length: 350
      Ratio: 5.134          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-662-128A-2 x AAD09280 ..

Align seg 1/1 to: AAD09280 from: 1 to: 4847

1 MetProLysLysArgLysValSerAsnLeuLeuThrValHisGlnAs 17
|||||
2775 ATGCCCAAGAAAGAGGAGGTGTCCAAATTACTGACCGTACACCAAA 2824
17 nLeuProAlaLeuProValAspAlaThrSerAspGluValArgLysAsnL 34
|||||
2825 TTTCCTGATTAACCGGTGATGCAACGAGTGAAGGTTGCAAGAACCC 2874
34 eumetAspmetPheArgAspArgGlnAlaPheSerGluHisThrTyrLys 50
|||||
2875 TGATGACATGTTTCAGGATCGCCAGGCGTTTCTGACCAATACCTGGAA 2924
51 MetLeuLeuSerValCysArgSerTyrPalaAlaTyrCysLysLeuAsnAs 67
|||||
2925 ATGCTTCTGTCCGTTTCCGCGGTGCGGTGAGCGGCGCATGTCACAGTTGAATTA 2974

```

```

67 narglystrpPheProAlaGluProGluAspValArgAspTyrLeuLeuT 84
   |||||
2975 CCGAATATGTTCCCGCAGAACCTGAAGATGTCGCGATTATCTTCTAT 3024
   |||||
84 yrlEuLnAlaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuGly 100
   |||||
3025 ATCTTAGCGCGCGGCTGTGGCAGTAAACATATCCAGCAACATTTGGGC 3074
   |||||
101 GlnLeuAsMetLeuHisArgArgSerGlyLeuProArgProSerAspSe 117
   |||||
3075 CAGCTAAACATGCTTATCGTCGTCGGGTCGCCACGACCAAGTGACG 3124
   |||||
117 rAsnAlaValSerLeuValMetArgArgIleArgLysGluAsnValAspA 134
   |||||
3125 CAATGCTGTTCACTGCTATGCGGGGATCCGAAAGAAACGTTGATG 3174
   |||||
134 lAgIyGluArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheAsp 150
   |||||
3175 CCGGTGAACGTGCAAAACAGGCTTAGCGTTGCAACGCACTGATTTGAC 3224
   |||||
151 GlnValArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAs 167
   |||||
3225 CAGGTTCGTTCACTGATGAAATAGCCATGCTCCAGGATATACGTPA 3274
   |||||
167 nLeuAlaPheLeuGlyIleAlaTyrAsnThrLeuLeuArgIleAlaGlu 184
   |||||
3275 TCTGGCATTTCTGGGATTTGCTATACACCTGTTACGTAAGCCGAA 3324
   |||||
184 lAlaArgIleArgValLysAspIleSerArgThrAspGlyArgMet 200
   |||||
3325 TTGGCAGATCAGGTTAAAGTATCTCAGCTACGACGAGTGAGAAATG 3374
   |||||
201 LeuIleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyValG 217
   |||||
3375 TTAATCCATATATGGCAGAACCAAAACGCTGGTTAGCACCGAGGTGA 3424
   |||||
217 uLysAlaLeuSerLeuGlyValThrLysLeuValGluArgTrpIleSerV 234
   |||||
3425 GAGGCACTTACCTGGGGGTAACCTGAACCTGTGAGAGAGGATTCGC 3474
   |||||
234 aLserGlyValAlaAspAspProAsnAsnTyrLeuPheCysArgValArg 250
   |||||
3475 TCTGTGGTGTAGTGTGATGATCCGAATACTACTGTTTGGCGGGTCA 3524
   |||||
251 LysAsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgA 267
   |||||
3525 AAAATATGGTGTGGCGGCCCATCTGCCACGACGCACTATCAACTCCGC 3574
   |||||
267 aLeuGluGlyIlePheGluAlaThrHisArgLeuIleTyrGlyAlaLysA 284
   |||||
3575 CCTGGAAGGGATTTTTCAGACCACTCATCGATTGATTACGGCGCTAAG 3624
   |||||
284 sPAspSerGlyGlnArgTyrLeuAlaTrpSerGlyHisSerAlaArgVal 300
   |||||
3625 ATGACTCTGTGCAGATACCTGCGCTGTGGACACACTGCCCGTCTC 3674
   |||||
301 GlnValAlaTrpAspMetAlaArgAlaGlyValSerIleProGluIleMe 317
   |||||
3675 GGAACCCCGCGAGATATGCGCGCGGTGAGTTCAATACCGAGATCAT 3724
   |||||
317 tGlnAlaGlyGlyTyrThrAsnValAsnIleValMetAsnTyrIleArgA 334
   |||||
3725 GCAAGCGTGTGGCTGACCAATGTAAATATTGTCATGAACTATATCCGTA 3774
   |||||
334 sNeuAspSerGluThrGlyAlaMetValArgLeuLeuGluAspGlyAsp 350
   |||||
3775 ACCGTGATTAATGAACAGGGGCAATGTGCGCTGCGAAGATGCGCAT 3824
   |||||

seq_name: /STD1/gcgdata/geneseq/geneseqn-embL/NA2001A.DAT:AAD09269
seq_documentation_block:
ID   AAD09269 standard: DNA: 4960 BP.

```

```

XX
AC   AAD09269;
XX
DT   12-SEP-2001 (first entry)
XX
DE   pcMV-I-Cre-pA vector DNA.
XX
KW   DNA recombinase domain; protein transduction domain; PTD;
KW   gene alteration; fusion protein; Human immunodeficiency virus;
KW   HIV; PCMV-I-Cre-pA vector; ds.
XX
OS   Unidentified.
XX
PN   WO200149832-A2.
XX
PD   12-JUL-2001.
XX
PE   05-JAN-2001; 2001MO-EP00060.
XX
PR   07-JAN-2000; 2000EP-0100351.
XX
PR   10-NOV-2000; 2000EP-0124595.
XX
PA   (ARTE-) ARTEMIS PHARM GMBH.
XX
PI   Schwenk F;
XX
DR   WPI; 2001-441873/47.
XX
PT   Using site-specific DNA recombinase domain/protein transduction domain
PT   fusion proteins for inducing target gene alterations in organisms or
PT   cell cultures -
XX
PS   Example: Page 62-63; 85pp; English.
XX
CC   The present invention relates to use of fusion proteins comprising
CC   a site-specific DNA recombinase domain e.g. Cre and a protein
CC   transduction domain (PTD) e.g. the Human immunodeficiency virus
CC   (HIV) derived TAR peptide, for preparing an agent for inducing
CC   target gene alterations in a living organism or cell culture. The
CC   present invention also provides a method for inducing gene
CC   alterations in living organisms using the fusion proteins of the
CC   invention. The present sequence is a pcMV-I-Cre-pA vector DNA.
XX
SQ   Sequence 4960 BP; 1225 A; 1213 C; 1296 G; 1226 T; 0 other:

alignment_scores:
      Quality: 1797.00      Length: 350
      Ratio: 5.134          Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-662-128A-2 x AAD09269 ..
Align seg 1/1 to: AAD09269 from: 1 to: 4960

1 MetProLysLysLysArgLysValSerAsnLeuLeuThrValHisGlnAs 17
   |||||
977 ATGCCCAAAAGAGAGAGAGAGTGTCCAAATTACTGACCGTACACCAAAA 1026
   |||||
17 nLeuProAlaLeuProValAspAlaThrSerAspGluValArgLysAsnL 34
   |||||
1027 TTTGCTGTGATTACCGGTGATGATGCAACGATGATGAGGTGCGCAAGAAC 1076
   |||||
34 eMetAspPhePheArgAspArgGlnAlaPheSerGluHisThrTrpLys 50
   |||||
1077 TCATGACATGTTCCAGGATGCGCCAGGCGTTTCTGAGCATACCTGGAAA 1126
   |||||
51 MetLeuLeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAs 67
   |||||
1127 ATGCTTCTGTCCGTTGCCGCGTGGCGGCGCATGTCGCAAGTTGAAATA 1176
   |||||
67 nArgLysTrpPheProAlaGluProGluAspValArgAspTyrLeuLeuT 84

```

```

1177 CCGGAAAGGTTTCCCGCAGACCTGAGATGCTGGCATATCTTCAT 1226
84 YLEUGLALARGGLYLEUALVALLYSTHRIEGLINGLNIHLSLEUGLY 100
1227 ATCTTCAGCGCCCGCGTGTGGCAGTAAACATATCCAGCAACATTTGGGC 1276
101 GINLEUASNMETLEUHSIARGTSERGLYLEUPROARGPROSERASPSE 117
1277 CAGCTAAACATGCTTCATGCTGCGGCTGCGCAGCAGCAAGTGACAG 1326
117 FASNALVALSERLEUVALMETARGARGLIARGLYSLUBSNVALASPA 134
1327 CAATCTGTTTCACGTGTTATCGCGCGCATCCGAAAAAGAAACGTGATG 1376
134 IAGLYLUARGALALYSGINALALEUALAPHEGLIARITHTASPHASP 150
1377 CCGGTGAACGTGCAGAAACAGGCTTAGCGTTGCAACGCACATGATTTGAC 1426
151 GINVALARGSERLEUMETGLUBASERSPARGCYSGINASPILEARGAS 167
1427 CAGGTTCGTTCACTCATGAAATAGCGATGCTGCCAGGATATACGTAA 1476
167 NLEUALAPHELEUGLYLEALATYRASNTHRELEUNARGYLEAGLUI 184
1477 TCTGCGATTTCTGGGAGTTGCTTATACACCCCTGTAGCTATAGCCGAAA 1526
184 IALALARGILEARVALLYSASPILIESERARGTHASPGLYLYARGMET 200
1527 TTGCCAGATCTTACGTTAAAGATATCTACGTACTAGCGTGAGGAATG 1576
201 LEULIEHISILEGLYARTHRYSRTHREUVALSERTHRALGLIYVALGI 217
1577 TTTATTCATATTTGGCAGAACGAAACGCTGTTAGCACCGCAGGTGAGA 1626
217 UULYSALALEUSERLEUGLYVALTHRLYSLEUVALGLUARGTPILESERV 234
1627 GAAGGCACATTAGCTGGGGGTAACCTGCTGCGAGCATGATGATTTCCG 1676
234 ALSERGLYVALALASPASPROMASNANTYRLEUPHECYSTRGYVALARG 250
1677 TCTCTGCTGAGCTGATGATCCGAATACCTGCTGTTTGGCGGCTACAA 1726
251 LYSASNGLYVALALALAPROSERALATHRSERGLINLEUSERTHRARGL 267
1727 AAAAATGGTGTTCGCCGCCCATCTGCCACACGACCTATCAACTGCGGC 1776
267 ALDEUGLGLYILEPHEGLUALATHRHSARGLEULIETRYRGYALALYSA 284
1777 CCTGGGAAGGATTTTGAAGCAACTCATGATTTGATTTACGGCGCTAAG 1826
284 SPASPSPERGLYGNARGTYRLEUALATRPSEGLYHISSEALAAARGVAL 300
1827 ATGACTCTGTGTCAGAGTACTGCTGCTGCTGACAGACAGTGCCGCTGTC 1876
301 GLYALALALARGASPMETALARGALAGLYVALSERILEPROGLUILEME 317
1877 GGAGCCCGCGAGATATAGCCCGCTGGAGTTTCAATACCGGAGATCAAT 1926
317 TGINALAGLYLYTRPHTASNVALASNILEVALMETASNTYRILEARGA 334
1927 GCAAGCTGCTGGCTGGACCAATGTAAATATTGTCAATGAACTAATATCCG 1976
334 SNLEUASPSEGLUTHTHGLYALAMETVALARGLEULEUGLUNASPGLYASP 350
1977 ACCTGGATAGTGAACAGGGGCAATGCTGCTGCGGAAGATGGCGAT 2026

```

seq_name: /STDS1/gcgdata/geneseq/geneseqn_emb1/NA2001A.DAT.AAD04928

seq_documentation_block:

ID AAD04928 standard; DNA; 5365 BP.
XX
AC AAD04928;

```

XX 17-JUL-2001 (first entry)
DE Retroviral vector PBABP-pgkCre used as gene trap in embryonic stem cells.
XX
XX Gene trapping construct; conditional mutation; unidirectional inversion;
XX recombinase recognition sequence; RRS; disruption cassette;
XX selection cassette; transgenic organism; retroviral vector;
XX PBABP-pgkCre; cre recombinase; phosphoglycerate kinase; PGK promoter; ds.
XX
OS Chimeric - Moloney murine leukemia virus.
OS Chimeric - Mus sp.
OS Chimeric - Bacteriophage P1.
OS Chimeric - Rhesus macaque polyoma virus.
XX
FH Key
FT Location/Qualifiers
FT LTR
FT 8..480
FT /tag= a
FT /note= "5' partial MMLV U3, MMLV R and MMLV U5"
FT /partial
FT 8..335
FT /tag= b
FT /note= "5' partial MMLV U3"
FT /partial
FT 336..402
FT /tag= c
FT /rpl_type= DIRECT
FT /note= "5' MMLV R direct repeat"
FT 403..480
FT /tag= d
FT /note= "5' MMLV U5"
FT 481..1374
FT /tag= e
FT /note= "MMLV primer binding site and extended packaging signal"
FT 1417..1921
FT /tag= f
FT /note= "Mouse phosphoglycerate kinase (PGK) promoter"
FT 1972..3024
FT /tag= g
FT /product= "Bacteriophage P1 Cre recombinase protein with a simian virus 40 large T-antigen N-terminal nuclear localisation signal"
FT 3088..3168
FT /tag= h
FT /note= "Promoter/enhancer deleted MMLV U3"
FT 3187..3332
FT LTR
FT /tag= i
FT /note= "3' MMLV R and MMLV U5"
FT /partial
FT 3187..3253
FT /tag= j
FT /rpl_type= DIRECT
FT /note= "3' MMLV R direct repeat"
FT 3254..3332
FT /tag= k
FT /note= "3' MMLV U5"
XX
XX WO200129208-A1.
XX
XX 26-APR-2001.
XX
XX 16-OCT-2000; 2000WO-EP10162.
XX
XX 16-OCT-1999; 99EP-0120592.
XX 27-OCT-1999; 99US-0162016.
XX
XX (ARTE-) ARTEMIS PHARM GMBH.
XX (FRAN-) FRANKEN BIOTECHNOLOGIE AG.
XX Kuehn R, von Melchner H, Altschmied J;
XX WPI; 2001-308486/32.
XX

```

XX New gene trapping construct capable of causing conditional mutations in
PT genes, comprises functional DNA segment inserted in sense or antisense
PT direction relative to gene to be trapped

PS Example 1; Page 55-57; 78bp; English.

XX The present invention relates to a conditional gene trapping construct
CC capable of causing conditional mutations in genes. The gene trapping
CC construct comprises two functional DNA segments, each being flanked by
CC two recombinase recognition sequences (RRSs) specific to site specific
CC standard DNA segment. One of the DNA segment (disruption cassette) is
CC inserted in antisense orientation relative to the transcriptional
CC orientation of the gene to be trapped. The other DNA segment (selection
CC cassette) is inserted in sense direction relative to the transcriptional
CC orientation of the gene to be trapped. The cell comprising the gene
CC trapping construct is useful for the identification and/or isolation of
CC genes. The transgenic organism comprising the gene trapping construct is
CC useful to study gene function at various developmental stages. The gene
CC trapping construct is useful for mutationally inactivating all cellular
CC genes. The present sequence is retroviral vector PBAB-pgkire which is
CC used as a gene trap in embryonic stem (ES) cells. The vector
CC PBAB-pgkire is constructed by inserting bacteriophage P1 cre recombinase
CC coding region with a simian virus 40 large T-antigen N-terminal nuclear
CC localisation signal and mouse phosphoglycerate kinase (PGK) promoter into
CC Moloney murine leukemia virus based vector PBAB-puro.

SQ Sequence 5365 BP; 1208 A; 1448 C; 1402 G; 1307 T; 0 other:

alignment_scores:

Quality: 1797.00 Length: 350
Ratio: 5.134 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-662-128A-2 x AAD04928

Align seg 1/1 to: AAD04928 from: 1 to: 5365

1 MetProLysLysArgLysValSerAsnLeuThrValHisGlnAs 17
1972 ATGCCCAAGAAAGAGAGAGAGTGTCCAAATTACTGACCGACCAAAA 2021
17 nleuProAlaLeuProValAlaThrSerAspGluValArgLysAsnL 34
2022 TTGGCTGCATTACCGGTGCGATGCAACGAGTGAAGTTCGCAAGAAC 2071
34 eumetAspMetPheArgAspArgGlnAlaPheSerGluHisThrTyrLys 50
2072 TGATGGACATGTTCAGGGATCGCCAGCGTTTCTGAGCTACTCTGGAAA 2121
51 MetLeuSerValCysArgSerTyrAlaAlaTyrCysLysLeuAsnAs 67
2122 ATGCTTCTGTCGTTTCCGCGTGTGGCGGCGCATGGTCAAGTTGAATA 2171
67 nArgLysTyrPheProAlaGluProGluAspValArgAspTyrLeuLeuT 84
2172 CCGGAATGTTGCCCGCAGAACCTGAAGATGTTCGCGATTAATCTCTAT 2221
84 YrLeuGlnAlaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuGly 100
2222 ATCTTCAAGGCGCGGCTGTGGCAGTAAACATATCCAGCAACATTTGGCG 2271
101 GlnLeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSe 117
2272 CAGCTAAACATGCTTCATCGTCCGCTCCGCGCTGCGACGACCAAGTGACG 2321
117 rAsnAlaValSerLeuValMetArgArgTleArgLysGluAsnValAspA 134
2322 CAATGCTGTTTCACTGTTATGCGGCGGATCCGAAAGAAAGAAACGTTGATG 2371

134 IagLGIARGLAlaLysGlnAlaLeuAlaPheGluArgThrAspPheAsp 150
2372 CCGGTGAACGTGCAAAACAGGCTCTACCGTTCACAGCGACATGTTTCGAC 2421
151 GlnValArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAs 167
2422 CAGGTGCTTCACATCATGCAAAATACGATCGCTGCCAGGATATACGTAA 2471
167 nleuAlaPheLeuGlyIleAlaTyrAsnThrLeuLeuArgGlyIleAlaGluI 184
2472 TCTGGCAATTTCTGGGATGCTTATTAACACCTGTACTGATATACCGAAA 2521
184 leuAlaArgIleArgValLysAspIleSerArgThrAspGlyLysArgMet 200
2522 TTGGCAGGATCAGCGTTAAAGATATCTACGACTGACGATGCGGAATG 2571
201 leuIleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyValG 217
2572 TTATCCATATTGGCAGAAACGAAACGCTGTACACCGCAGGTGTGA 2621
217 ulysAlaLeuSerLeuGlyValThrLysLeuValGluArgTyrPheSerV 234
2622 GAAGGCACTTAGCTGGGGGTAACTAACTGCTGCGAGGATGATTTCCG 2671
234 alSerGlyValAlaAspAspProAsnAsnTyrLeuPheCysArgValArg 250
2672 TCTCTGTGTACTGATGATGATGCAATTAACCTGTTTGGCGGGTCA 2721
251 LysAsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAl 267
2722 AAAAATGCTGTGCGCGCCCATCTGCGCACGACGACGATCACTACGCGC 2771
267 aleuGluGlyIlePheGluAlaThrHisArgLeuIleTyrGlyAlaLysA 284
2772 CCTGAGAGGATTTTGAAGCACTCATCGATGATTTAGCGCGCTAAG 2821
284 sPAspSerGlyGlnArgTyrLeuAlaTyrPheSerGlyHisSerAlaArgVal 300
2822 ATGACTCTGCTGAGAGATACCTGCTGTGTGGACACAGCGCCGCTGTC 2871
301 GlyAlaAlaArgAspMetAlaArgAlaGlyValSerIleProGluIleMe 317
2872 GGAGCGCGCGAGATATGCGCGCGCTGAGTTTCATATACCGAGATCAT 2921
317 tGlnAlaGlyIleTyrPheAsnValAsnIleValMetAsnTyrIleArgA 334
2922 GCAAGCTGTGTGCGGACCAATGTAATATTTGCATGCACTATATCCGTA 2971
334 snLeuAspSerGluThrGlyAlaMetValArgLeuGluGluAspGlyAsp 350
2972 ACCTGAGTATGTAACAGAGGGCAATGTGCGCTGCGAAGATGGCGAT 3021
seq_name: /STDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.AAX19900
seq_documentation_block:
ID AAX19900 standard; DNA; 2293 BP.
XX AAX19900:
XX
XX
XX 11-JUN-1999 (first entry)
XX
XX DE Plasmid pOG304M containing mPl promoter and Cre coding sequences.
XX
XX KW Mouse; mPl; promoter; site-specific recombination; Cre recombinase;
XX
XX KM proteasine 1; fusion gene; transgenic mouse; ss.
XX
XX OS Synthetic.
XX
XX OS Mus musculus.
XX
XX PN W09910488-A1.
XX
XX PD 04-MAR-1999.
XX

```

PF      28-AUG-1998;   98WO-US1785Z.
XX
XX      28-AUG-1997;   97US-0919501.
PR
XX      (SALK ) SALK INST BIOLOGICAL STUDIES.
PA
XX      O'Gorman S,  Wahl G;
XX      WPI; 1999-190617/16.
XX
XX      New nucleic acid construct, comprising a germline-specific promoter
PT      operably linked to a recombinase coding sequence - useful for
PT      producing subtle, conditional or tissue-specific mutations
XX
XX      Example 1; Page 48; 54pp; English.
XX
XX      The present invention describes a nucleic acid construct, comprising a
CC      germline-specific promoter operably linked to a recombinase coding
CC      sequence. The construct can be useful for modulating agronomic
CC      characteristics, including (a) biotic stress tolerance, increased
CC      resistance to herbicides, pest damage, and viral, bacterial, and fungal
CC      diseases, improvement of crop nutritional quality, reduction of
CC      post-harvest losses, improved quantity and composition of endogenous
CC      properties, and production of new plant-derived compounds. The
CC      production of alleles containing single recombinase target sites and a
CC      desired mutation are useful for producing subtle and conditional
CC      alterations that require generation of alleles with minimal structural
CC      alterations. They are also useful for generating null alleles.
CC      mpt-mediated nucleic acid constructs are useful for producing embryos
CC      containing genetically lethal alleles. The recombinase constructs are
CC      expressed at high levels in the germ line, but not to a functionally
CC      significant degree in ES cells or embryonic or adult somatic tissue.
CC      This enables modulation of recombined target nucleic acid sequences in
CC      the early embryo. In addition, tissue-specific and conditional
CC      recombinase tools permit the generation of transgenic plants and
CC      animals. The present sequence represents a plasmid containing mouse
CC      protamine 1 promoter and Cre coding sequences from an example of the
CC      present invention.
XX
XX      Sequence 2293 BP; 555 A; 559 C; 603 G; 576 T; 0 other;
SQ

```

alignment_scores:

	Quality: 1792.00	Length: 349
Ratio: 5.135	Gaps: 0	
Percent Similarity: 100.000	Percent Identity: 100.000	

alignment_block:

```

US-09-662-128A-2 x AAX19900 ..
Align seg 1/1 to: AAX19900 from: 1 to: 2293

```

```

2      ProlylsylsArGlyValSerAsnLeuThrValHisGlnAsnLe 18
      |||||||
699    CCCAAGAGAGAGAGAGAGGTTCCATTACGACCGTACACCAAAATTT 748
      |||||||
18      uproAlaLeuProValAspAlaThrSerAspGluValArgLysAsnLeuM 35
      |||||||
749    GCGTCGATACCGTCGATCAACGAGTGTGAGGTTCCGCAAGAACCTGA 798
      |||||||
35      eLaspMetPheArgAspArgGlnAlaPheSerGluHisThrTrpLysMet 51
      |||||||
799    TGGACATGTTCCAGGATGCCAGGCGTTTCTAGACATACCTGGAAATAG 848
      |||||||
52      LeuLeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAsn 68
      |||||||
849    CTTCGTCCGTTTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 898
      |||||||
68      gLysTrpPheProAlaGluProGluAspValArgAspTrpLysLeuTrp 85
      |||||||
899    GAAATGGTTTCCCGCAGAACCTGAAGATGTTCCGATTAATCTTCTATATC 948
      |||||||
85      euGlnAlaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuGlyGln 101

```

```

|||||
949    TTAGAGCGCGCGGTCTGCAGATAAAACATACCAACACATTGGCCAG 998
|||||
102    LeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAs 118
|||||
999    CTAACATGCTTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1048
|||||
118    nAlaValSerLeuValMetArgArgIleArgLysGluAsnValAspAlaG 135
|||||
1049    TGCTGTTTCACGTGTTATGCGGGATCCGAAAGAAACGTTGATGCCG 1098
|||||
135    LysIuArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheAspGln 151
|||||
1099    GTGAACGTCGCAAAACAGGCTCTAGCGTTGAAACGCACTGATTTCCAGC 1148
|||||
152    ValArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAsnLe 168
|||||
1149    GTTCGTTCTACTCATGTAATAATACGATCGCTGCAGAGATATACGTATCT 1198
|||||
168    uAlaPheLeuGlyIleAlaIleArgAsnThrLeuLeuArgIleAlaGluIle 185
|||||
1199    GGCATTTCTGGGAGATGCTTATTAACACCTGTACGTATATAGCGAAATTG 1248
|||||
185    IaArgIleArgValLysAspIleSerArgThrAspGlyArgMetLeu 201
|||||
1249    CCAGATCAGGGTTAAAGATATCTCAGCTACTGACGCGTGGAATAATGTTA 1298
|||||
202    IleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyIleGly 218
|||||
1299    ATCCATTTGGCAGACAGAAACCTGTTAGCACCCAGCTGTACAGAA 1348
|||||
218    sAlaLeuSerLeuGlyValIleThrLysLeuValGluArgTrpIleSerVal 235
|||||
1349    GGCACCTTGCCCTGGGGGTAACTAAACGTGTGACCGATGATTCGTCGT 1398
|||||
235    eGCIValAlaAspAspProAsnAsnTrpLeuPheCysArgValArgLys 251
|||||
1399    CTGGTGTCGTCGATGATCGAATACCTACCTGTTTCCGCGGTACAGAAA 1448
|||||
252    AsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAlaLe 268
|||||
1449    AATGGTGTGGCCGCCCATCTGCCACAGCCAGCTATCACTCGGCCCT 1498
|||||
268    uGluGlyIlePheGlnAlaThrHisArgLeuIleArgLysAlaLysAsp 285
|||||
1499    GGAAGGATTTTGAAGCAACTCATGATGATTACGCGCCTTAAGGATG 1548
|||||
285    spSerGlyLysArgGlyLeuAlaTrpSerGlyHisSerAlaArgValGly 301
|||||
1549    ACTCTGTGTCAGAGATACCTGCGCTGTCGTCGTCGTCGTCGTCGTCG 1598
|||||
302    AlaAlaArgAspMetAlaArgAlaGlyValSerIleProGluIleMetG 318
|||||
1599    GCCGCCCGAGATATGCGCCGCGCTGAGTTTCATACCGAGATCATCA 1648
|||||
318    nAlaGlyGlyTrpThrAsnValAsnIleValMetAsnTrpIleArgAsnL 335
|||||
1649    AGCTGGTGCTGGACCAATGTAATATTTCTCATGAATCATATATCCGTA 1698
|||||
335    euAspSerGluThrGlyAlaMetValArgLeuLeuGluAspGlyAsp 350
|||||
1699    TGGATGATGAAACAGGCGCATGTGCGCTGCTGGAAGATGCGCAT 1745
|||||
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/AA2001A.DMT:AAH26331
seq_documentation_block:
ID      AAH26331 standard; DNA; 5261 BP.
XX
XX      AAH26331:
AC
XX      02-OCT-2001 (first entry)
DT
XX      Recombinant adenovirus delta-E1 Cre-Loxp Ad.
DE

```

```

XX Adenovirus; delta-E1 Cre-LoxP Ad; adeno-associated virus; vector;
KW gene therapy; ds.
XX
OS Chimeric - Mastadenovirus.
OS Chimeric - Human cytomegalovirus.
OS Chimeric - Rhesus macaque polyoma virus.
XX
FH
FT Key Location/Qualifiers
FT promoter 551..1186
FT /tag= a
FT /note= "CMV promoter"
FT polyA_site 2251..2476
FT /tag= b
FT /note= "SV40 polyA site"
FT m1sc_recomb 2476..2520
FT /tag= c
FT /function= "loxP site"
XX
XX MO200155361-A2.
XX
XX 02-AUG-2001.
XX
XX 26-JAN-2001; 2001WO-US02709.
XX
XX 26-JAN-2000; 2000US-0178536.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Hardy SF;
XX
XX WPI; 2001-483239/52.
XX
XX
XX Producing recombinant adeno-associated virus (rAAV) vector, by stably
XX infecting eukaryotic host cell with rAAV vectors, helper adenovirus or
XX herpes virus, recombinant adenovirus or herpes vectors -
XX
XX Disclosure; Page 60-62; 63pp: English.
XX
XX
XX The present sequence is that of recombinant adenovirus delta-E1
XX Cre-LoxP Ad. In which the E1 gene of the wild-type adenovirus is
XX substituted by plasmid DNA comprising a cytomegalovirus (CMV)
XX promoter inserted immediately downstream of the CMV promoter
XX followed by a sequence from SV40 that specifies polyadenylation in
XX mDNA, and a loxP site. The recombinant adenovirus is used as an
XX induction system in methods of the invention. Thus, methods
XX and compositions are provided for producing recombinant adeno
XX associated virus (rAAV) vector particles by: (a) introducing into
XX a host cell (1) AAV packaging plasmid pflloxAAV (see AAH26326), (11)
XX a recombinant viral vector encoding plasmid, and (111) a plasmid
XX encoding herpesvirus, cytomegalovirus or adenoviral functions, or
XX a herpesvirus, cytomegalovirus or adenovirus itself, to produce
XX CC flox AAV particles and rAAV particles; and (b) introducing into a
XX second host cell (1) the rAAV particles or (a), (11) a vector
XX that directs expression of Cre, and (11) a vector which directs
XX expression of herpes virus, cytomegalovirus or adenovirus helper
XX functions, such that rAAV vector particles are produced. The
XX vectors are useful for in vivo or in vitro gene therapy and also
XX for in vitro recombinant protein production.
XX
XX Sequence 5261 BP; 1309 A; 1238 C; 1329 G; 1385 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 1792.00      Length: 349
XX      Ratio: 5.135      Gaps: 0
XX      Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX alignment block:
XX      US-09-662-128A-2 x AAH26331      ..
XX
XX Align seg 1/1 to: AAH26331 from: 1 to: 5261

```

```

2 ProLysLysLysArgLysValSerAsnLeuThrValHisGlnAsnLe 18
1201 CCAAAGAGAGAGAGAAAGTTTGCATTTACTGACCGTACACCAAAATTT 1250
18 uProAlaLeuProValAspAlaThrSerAspGluValArgLysAsnLeu 35
1251 GCCTGCATTACCGCTGATGCACAGAGTATGAGTTCGCAAGAACCTGA 1300
35 eAspMetPheArgAspArgGlnAlaPheSerGlnHisThrLysMet 51
1301 TGGACATGTTCCAGAGATCCCGACGCTTTTCTGACATACATCGGAAATG 1350
52 LeuLeuSerValCysArgSerTrpAlaLysValArgAspLysLeuAsnArg 68
1351 CTTCGTCCGTTGGCCGTCGTGGCGGCATGTGCAGATTGAATACCG 1400
68 GlySerTrpPheProAlaGluProGluAspValArgAspLysLeuLys 85
1401 GAATGCTTTCCCGCGAAGCTGAAAGATGTTCGCAATATCTCTATATTC 1450
85 euGlnAlaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuGlyGln 101
1451 TTCAGCGCGCGCTGCGCAGTAAACTATCCAGCAACATTTGGGCCAG 1500
102 LeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAs 118
1501 CTAAACATGCTTCATCGTCGCGCTCCGCGCTCCACAGACAGTGCACCAA 1550
118 nAlaValSerLeuValMetArgArgIleArgLysGlnAsnValAspAla 135
1551 TCCTGTTTCACATGCTTATGCGCGCGCATCCGAAAGAAACGTTGATCCG 1600
1351 LysGlnAlaArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheAspGln 151
1601 GTAAACATGCTTCATCGTCGCGCTCCGCGCTCCACAGACAGTGCACCAA 1650
152 ValArgSerLeuMetGluAsnSerAspArgCysGlnAspLysLeuArg 168
1651 GTTCGTTACATCATGGAATAAGCGATCGCTCCAGGATATACGTAAATCT 1700
168 uAlaPheLeuGlyIleAlaLysAsnThrLeuLeuArgIleAlaGlyIle 185
1701 GGCATTCTCGGGAATTCCTTATACACCGTTCAGTATGCGGAATTCG 1750
1851 LaArgIleArgValLysAspLysSerArgThrAspGlyArgMetLeu 201
1751 CCAGGATCAGGGTTAAAGATATCTCACGTACTGACGTTGGAGAAATGTTA 1800
202 IleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyValGly 218
1801 ATCCATATTGGCAGAAACGAAACCGCTGTTAGCACCGCAGGTGTAGAGAA 1850
218 sAlaLeuSerLeuGlyValThrLysLeuValGluArgTrpLysSerVal 235
1851 GGCCTTAAGCCTGGGGTTAACTAACTGGTCGACGATGATTCGCTCT 1900
235 eArgLysValAlaAspProAsnAsnThrLeuPheGlyArgValArgLys 251
1901 CTGGTGAAGTGAATGATCCGAATATCTACCTTTTGGCGGGTCCAGAAA 1950
252 AsnGlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAla 268
1951 AATGGTGTGCGCGCCCATCTGCCACACCGCATATCAACTCCGGCCCT 2000
268 uGluGlyIlePheGluAlaThrHisArgLeuLysLeuValAlaLysAsp 285
2001 GGAAGGATTTTGAAGCAATCATCGATGATTTACCGCCCTAAAGATG 2050
285 sPserGlyLysIleArgLysLeuAlaTrpSerGlyHisSerAlaArgValGly 301
2051 ACTGTGTCAGAGATACCTGGCTGTGTCGACACAGTCCGCTGTGCGA 2100
302 AlaAlaArgAspMetAlaArgAlaGlyValSerIleProGluIleMetGln 318

```

```

|||||
2101 GCGCGGAGATATGCGCCGCTGAGTTTCATACCGGAGATCATGCA 2150
318 nalaiglyglttrpThrAsnValAsnIleValMetAsnTrpIleArgAsnL 335
2151 AGCTGGTGGCTGAGCAATGTAATTTCTCATGACTATATCCGTAAAC 2200
335 euAspSerGluTrpGlyAlaMetValArgLeuLeuGluAspGlyAsp 350
2201 TGGATATGTGAACAGGGGCAATGTGCGCTGCTGGAAGATGCGCAT 2247

seq_name: /SID1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAD09257

seq_documentation_block:
ID AAD09257 standard; DNA; 1074 BP.
XX
AC AAD09257;
XX
DT 12-SEP-2001 (first entry)
XX
DE TAT-Cre fusion DNA.
XX
KW DNA recombinase domain; protein transduction domain; PTD;
KW gene alteration; TAT-Cre fusion protein; Human immunodeficiency virus;
KW HIV; ds.
XX
OS Chimeric - Human immunodeficiency virus.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..1074
FT /tag= a
FT /product= "TAT-Cre fusion protein"
XX
XX WO200149832-A2.
XX
XX 12-JUL-2001.
XX
XX PD 05-JAN-2001; 2001WO-EP00060.
XX
XX PF 07-JAN-2000; 2000EP-0100351.
XX
XX PR 10-NOV-2000; 2000EP-0124595.
XX
XX PA (ARTE-) ARTEMIS PHARM GMBH.
XX
XX PI Schwenk F.
XX
XX DR WPI: 2001-441873/47.
XX
XX P-PSDB: AAE05264.
XX
XX PT Using site-specific DNA recombinase domain/protein transduction domain
XX fusion proteins for inducing target gene alterations in organisms or
XX cell cultures -
XX
XX PS Claim 17; Page 26-27; 85pp; English.
XX
XX CC The present invention relates to use of fusion proteins comprising
XX a site-specific DNA recombinase domain e.g. Cre and a protein
XX transduction domain (PTD) e.g. the Human immunodeficiency virus
XX (HIV) derived TAT peptide, for preparing an agent for inducing
XX target gene alterations in a living organism or cell culture. The
XX present invention also provides a method for inducing gene
XX alterations in living organisms using the fusion proteins of the
XX invention. The present sequence is a DNA encoding TAT-Cre fusion
XX protein. The TAT sequence is from HIV.
XX
SQ Sequence 1074 BP; 268 A; 254 C; 299 G; 253 T; 0 other;

```

alignment_scores: Length: 348
Quality: 1765.00
Ratio: 5.086
Percent Similarity: 99.713 Percent Identity: 98.563

```

alignment_block:
US-09-662-128A-2 x AAD09257
..
Align seq 1/1 to: AAD09257 from: 1 to: 1074
3 LysLeuLysArgLysValSerAsnLeuThrValHisGlnAsnLeuP 19
28 CAACGGCGGGCGGCGATGCCAATTTACTGACCGTACACCAAAATTTGCC 77
19 oAlaLeuProValAspAlaThrSerAspGluValArgLysAsnLeuMe 36
78 TGCATTACCGCGTGTGATGCAACGAGTGATGAGTTTGCAGAAACCTGAG 127
36 spMetPheArgAspArgGlnAlaPheSerGluHisThrTrpLysMetLeu 52
128 ACATGTTCAAGGATCCGCCAGCGCTTTCTGAGCATPACTGAAATTCCTT 177
53 LeuSerValCysArgSerTrpAlaIleTrpCysLysLeuAsnArgLys 69
178 CTGTCCTTGCCGCGTGTGGCGGCATGTGCAAGTTGAATTAACCGGGA 227
69 strPheProAlaGluProGluAspValArgAspTrpLeuLeuTrpLeu 86
228 ATGGTTTCCCGCAGAACCTTAAGATGTTCCGCAATTATCTTATATCTTC 277
86 LnaIaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuGlnLeu 102
278 AGCGCGCGGTCTGCGCAATAAATATCCAGCAACATTTGGCCAGCTTA 327
103 AsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAsn 119
328 AACATGCTTCATCGTCCGCTCCGGCTCCACGACCAAGTACAGCAATCG 377
119 aValSerLeuValMetArgArgIleArgLysGluAsnValAspAlaGly 136
378 TGTTCACCTGGTATGCGCGCGATCCGAAAGAAAGATTGATGCCGGTG 427
136 LnaArgAlaLysGlnAlaLeuAlaPheGluArgTrpAspPheAspGln 152
428 AACGTGCAAAACAGGCTCTAGCTTCGACGCACTGATTCGACAGAGTT 477
153 ArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAsnLeu 169
478 GGTTCACCATGCAAAATATGAGATCGCTCCGAGATATGATATCTGCG 527
169 aPheLeuGlyIleAlaIleArgAsnThrLeuLeuArgIleAlaGluIle 186
528 ATTTCGGGATGCTTATTAACACCCCTGTACGATATGCAATTCGCA 577
186 rGluLeuValLysAspIleSerArgTrpAspGlyArgMetLeuIle 202
578 GGTACGGGTAAAGATATCTCAGTACTGACGGGAGCAATGTATATC 627
203 HisIleGlyArgThrLysThrLeuValSerThrAlaGlyValGluVal 219
628 CATATTGGCAGAACGAAACGCTGTGTGACCCGAGGTGAGCAAGACG 677
219 aLeuSerLeuGlyValThrLysLeuValGluArgTrpLysSerValSer 236
678 ACTTACCTCGGGGCTTACTAACTGTGAGGAGATTTCCGTCCTG 727
236 LysValAlaAspAspProAsnAsnTrpLeuPheCysArgValArgLys 252
728 GTGTAGCTGATGATCCGAATTAACCTCTTTTGGCGGCTGAGAAATAT 777
253 GlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAlaLeu 269
778 GGTGTGCGCGCCATCTGACACCGACCGACTATCACTCGCGCCCTGGA 827
269 uGlyIlePheGluAlaThrHisArgLeuIleTrpGlyAlaLysAspAsp 286
828 AGGATTTTGAAGCAACTCATGATGATTTAAGCGCGCTAAGATGACT 877

```



```

253 GlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAlaLeuG1 269
    |||
919 GGTGTGGCGCGCATCTGCACACAGCTATCAACTGCCGCCCTGGA 968
269 uGlyTLeuHeGluAlaThrHisArgLeuIleTyrGlyAlaAlaAspAspS 286
    |||
969 AGGATTTTGAAGCACTCATCTCATTCATTATACGGCGCTAAGATGACT 1018
286 erGlyGlnArgTyrLeuAlaTrpSerGlyHisSerAlaArgValGlyAla 302
    |||
1019 CTGGTCAGAGATACCTGCTGCTGTGACACAGTGCCTGTGCGAGCC 1068
303 AlaArgAspMetAlaArgAlaGlyValSerIleProGluIleMetGlnAl 319
    |||
1069 GCGCAGATATGCGCCGCGCTGAGTTCAATACCGAGATCATGCAGCC 1118
319 aGlyGlyTyrThrAsnValAsnIleValMetAsnTyrIleArgAsnLeuA 336
    |||
1119 TCGTGGCTGGACCATGTAATATGTGTCATGACTATATCCGTAACTGG 1168
336 sPSeGluThrGlyAlaMetValArgLeuLeuGluAspGlyAsp 350
    |||
1169 ATAGTGAACACAGCGCAATGTGCGCTGTGAGAGATGCGCAT 1212
seq_name: /SIND1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: AAD09267

seq_documentation_block:
ID AAD09267 standard; DNA: 1125 BP.
XX.
AC. AAD09267:
XX
DE 12-SEP-2001 (first entry)
XX
DE TATCrestreptag fusion DNA.
XX
XX DNA recombinase domain; protein transduction domain; PTD;
XX TATCrestreptag fusion protein; Human immunodeficiency virus; HIV;
XX gene alteration; ds.
XX
OS Chimeric - Human immunodeficiency virus.
XX
XX Chimeric - Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 1..1122
XX FT /*tag= a
XX FT /product= "TATCrestreptag fusion protein"
XX
XX MO200149832-A2.
XX
XX PD 12-JUL-2001.
XX
XX PF 05-JAN-2001; 2001WO-EP00060.
XX
XX PR 07-JAN-2000; 2000EP-0100351.
XX PR 10-NOV-2000; 2000EP-0124595.
XX
XX PA (ARTE-) ARTEMIS PHARM GMBH.
XX
XX PI Schwenk F;
XX DR WPI: 2001-441873/47.
XX DR P-PSDB: AAE05272.
XX
XX Using site-specific DNA recombinase domain/protein transduction domain
XX PT fusion proteins for inducing target gene alterations in organisms or
XX PT cell cultures -
XX
XX Claim 17; Page 52-54; 85pp; English.
XX
XX The present invention relates to use of fusion proteins comprising
XX a site-specific DNA recombinase domain e.g. Cre and a protein
XX transduction domain (PTD) e.g. the Human immunodeficiency virus
XX (HIV) derived TAT peptide, for preparing an agent for inducing

```

CC target gene alterations in a living organism or cell culture. The
 CC present invention also provides a method for inducing gene
 CC alterations in living organisms using the fusion proteins of the
 CC invention. The present sequence is a DNA encoding TATCrestreptag
 CC fusion protein. The TAT sequence is from HIV.

Sequence 1125 BP; 278 A; 265 C; 316 G; 266 T; 0 other;

alignment_scores:

Quality:	1761.00	Length:	348
Ratio:	5.075	Gaps:	0
Percent Similarity:	99.713	Percent Identity:	98.276

alignment_block:
 US-09-662-128A-2 x AAD09267

Align seg 1/1 to: AAD09267 from: 1 to: 1125

```

3 LysLysLysArgLysValSerAsnLeuThrValHisGlnAsnLeuP 19
    |||
28 CAACGCCGCCGCGCATGTCCTCAATTTACTGACCGTACACCAAAATTTGCC 77
19 GAlaLeuProValAspAlaThrSerAspGluValArgLysAsnLeuMetA 36
    |||
78 TGCATTACCGGTGATGCAACGAGTGTAGTGTCCCAAGAACCTGATGG 127
36 sPMeTPhaArgAspArgGlnAlaPheSerGluHisThrTyrLysMetLeu 52
    |||
128 ACATGTTCAAGGATCGCCAGCGCTTCTGAGCATACCTGGAAGAGCTT 177
53 LeuSerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnArgLys 69
    |||
178 CTGTCCGTTTCCGCTCGTGGCGGCGATGTGTCAAGTTGATTAACCGGA 227
69 sTrpPheProAlaGluProGluAspValArgAspTyrLeuLeuTyrLeuG 86
    |||
228 ATGGTTTCCCGCAGACACTGTAAGATGTTCCGAGTATCTCTATATCTTC 277
86 LnaIaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuGlnLeu 102
    |||
278 AGGCGCGCGGTCTGACAGTAAACCTATCCAGCAACATTTGGGCGACTA 327
103 AsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAsnAl 119
    |||
328 AACATCTTCAATCTGTCGCTCGCGGCTGCACAGCAAGTGCACACATATG 377
119 aValSerLeuValMetArgArgIleArgLysGluAsnValAspAlaGly 136
    |||
378 TGTTCACCTGCTTATGCGCGGATCCGAAAGAAAGCTTGATGCCGGTG 427
136 LArgAlaLysGlnAlaLeuAlaPheGluArgTyrAspPheAspGlnVal 152
    |||
428 AACGTCAAAACAGCTCTAGCGTTCGAAAGCCACTGATTTGCACAGTT 477
153 ArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAsnLeuAl 169
    |||
478 CGTTCACATGAGAAATAGCATGCTGCCAGATATACCTAATCTGGC 527
169 aPheLeuGlyIleAlaTyrAsnThrLeuLeuArgIleAlaGluIleAla 186
    |||
528 ATTTCCTGGGATGCTTATACACCCCTGTAGCGTATACCAAAATGTGCA 577
186 rGlyLeuValLysAspIleSerArgTyrAspGlyIleArgMetLeuIle 202
    |||
578 GGATCAGGTTAAGATATCTACCTACTGACGAGTGGAGATGTATATC 627
203 HisIleGlyArgThrLysThrLeuValSerThrAlaGlyAlaGluLysAl 219
    |||
628 CATATGGCAGAAACGAAACGCTGTTAGCACCGCAGGTGTAGAGAAAGC 677
219 aLeuSerLeuGlyValThrLysLeuValGluArgTrpIleSerValSerG 236
    |||

```



```

226 1YValAlaAspAspProAsnAsnTyrLeuPheCysArgValAlaArgLysAsn 252
      |||||||
574 GTGATAGCGATGATCCGAAATACACTGTTTGGCGGGTCAGAAAAAAT 525
253 GLYValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAlaLeuGI 269
      |||||||
524 GGTGTGGCGCGCATCTGCCACACGACCTATCAACTGCGCCCTGGGA 475
269 uGIYLePheGlnAlaThrHisArgLeuIleTyrGlyAlaLysAspAsps 286
      |||||||
474 AGGGATTTTGAAGCACTCATCATTTAGTTACGGCGCAAGAGATGACT 425
286 eRGlyGlnArgTyrLeuAlaTrpSerGlyHisSerAlaArgValGlyAla 302
      |||||||
424 CTGGTCAGAGATACCTGCTGCTGTGGACACACTGCCGCTGGAGACC 375
303 AlaArgAspMetAlaArgAlaGlyValSerIleProGlnIleMetGlnAl 319
      |||||||
374 GCGCAGATATGCGCCGCGCTGGAGTTTCATACCGAGATCATGCAAGC 325
319 aGIYGLYTrpThrAsnValAlaAsnIleValMetAsnTyrIleArgAsnLeuA 336
      |||||||
324 TGTGTGGCTGGACCAATGTAATATGTGTCATGAACTATACGTAACCTGG 275
336 sPSerGIuThrGlyAlaMetValAlaArgLeuLeuGluAspGlyAsp 350
      |||||||
274 ATAGTGAAACAGGGGCAATGTGTGCGCTGTGGAAGATGGCGAT 231

```

seq_name: /SID51/gcgdata/geneseq/geneseq - emb1/NA2001A.DAT.AAD09263

seq_documentation_block:

ID AAD09263 standard; DNA; 1623 BP.

AC AAD09263;

DT 12-SEP-2001 (first entry)

DE Delta VP22Cre-StrepTag fusion DNA.

KM DNA recombinase domain; protein transduction domain; PTD;

KW gene alteration; delta VP22Cre-StrepTag fusion protein;

KW Human immunodeficiency virus; HIV; Human spumaretrovirus; HSV; ds.

XX Chimeric - Human spumaretrovirus.

OS Chimeric - unidentified.

FH Key Location/Qualifiers

FT CDS 1..1620

FT /*tag= a

FT /product= "Delta VP22Cre-StrepTag fusion protein"

XX WO200149832-A2.

XX 12-JUL-2001.

XX 05-JAN-2001; 2001WO-EP00060.

XX 07-JAN-2000; 2000EP-0100351.

XX 10-NOV-2000; 2000EP-0124595.

XX (ARTE-) ARTEMIS PHARM GMBH.

XX PI Schwenk F;

XX WPI; 2001-441873/47.

XX P-PDB; AAE05270.

PT Using site-specific DNA recombinase domain/protein transduction domain
 PT fusion proteins for inducing target gene alterations in organisms or
 PT cell cultures -

PS Claim 17; Page 44-46; 85pp; English.

```

XX The present invention relates to use of fusion proteins comprising
CC a site-specific DNA recombinase domain e.g. Cre and a protein
CC transduction domain (PTD) e.g. the human immunodeficiency virus
CC (HIV) derived Tat peptide, for preparing an agent for inducing
CC target gene alterations in a living organism or cell culture. The
CC present invention also provides a method for inducing gene
CC alterations in living organisms using the fusion proteins of the
CC invention. The present sequence is a DNA encoding delta
CC VP22Cre-StrepTag fusion protein. The VP22 sequence is from
CC Human spumaretrovirus (HSV).
XX
SQ Sequence 1623 BP; 368 A; 440 C; 471 G; 344 T; 0 other:

```

```

alignment_scores:
      quality: 1759.00      length: 344
      Ratio: 5.113
Percent Similarity: 100.000      Percent Identity: 99.419

```

alignment_block:

US-09-662-128A-2 x AAD09263 ..

Align seg 1/1 to: AAD09263 from: 1 to: 1623

```

7 1YValAlaSerAsnLeuLeuThrValHisGlnAsnLeuProAlaLeuProVa 23
      :::::::::::
538 CGCATGTCCAAATTTACTGACCGTACACCAAAATTTGGCTGCATTACCGGT 587
23 1AspAlaThrSerAspGluValAlaArgLysAsnLeuMetAspMetPheArg 40
      |||||||
588 CGATGCACAGAGTATGAGTTCCGACAGAACCTGATGACATGTTCAGG 637
40 sPArgGlnAlaPheSerGlnHisThrTrpLysMetLeuLeuSerValCys 56
      |||||||
638 ATCGCCAGGCGTTTCTGAGCATACCTGGAATAATGCTCTGTCCGTTCG 687
57 ArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArgLysTrpPheProAl 73
      |||||||
688 CGGTCTGGCGCGCATGTGTCAAGTTGAATACCGGAATGTGTTCCCGC 737
73 aGIuPProGluAspValAlaArgAspTyrLeuLeuTyrLeuGlnAlaArgLyl 90
      |||||||
738 AGAACCTGAAGATGTTCCGATTAATCTCTAATCTTCAGCGCGCGTTC 787
90 eValAlaValLysThrIleGlnGlnHisLeuGlyGlnLeuAsnMetLeuHis 106
      |||||||
788 TGSCAGTAAAACTATCCAGCAACATTTGGGCCACCTAAACATGCTTCAT 837
107 ArgArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuVa 123
      |||||||
838 CGTCGGTCCGGGCTGCCACGACCAAGTACAGCATGCTGTTTCACTGCT 887
123 1MetArgArgIleArgLysGluAsnValAspAlaGlyLysArgAlaLysG 140
      |||||||
888 TATGCGGCGGATCCGAAGAAAGACGTTGATCCGCTGAACGTGCAAAAC 937
140 1AlaIleuAlaPheGlnuArgThrAspPheAspGlnValArgSerLeuMet 156
      |||||||
938 AGGCTTAGCGTTTCGAACGACGATGATTCGACCAGGTTGCTTCACTATG 987
157 GluAsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyI 173
      |||||||
988 GAAATATGCGATGCTGCCAGATATAGCTAATGTGGCATTTCTGGGAT 1037
173 eAlaTyrAsnThrLeuLeuArgIleAlaGluIleAlaArgIleArgVal 190
      |||||||
1038 TGCTTATACACCCCTGTATACGTAACCGAAATGTGCAGAGTACAGGTTA 1087
190 yAspIleSerArgThrAspGlyGlyArgMetLeuLeuHisIleGlyArg 206
      |||||||
1088 AAGATATCTACGTAAGTACGAGGTGGAGAAATGTTAATTCATATGGCAGA 1137

```

```

207 ThrLysThrLeuValSerThrAlaGlyValGluLysAlaLeuSerLeuG1 223
1138 ACGAAAAAGCTGGTTAGACACCGAGGTGTAGAGAAAGCCATTAGCCTGGG 1187
223 YValThrLysLeuValGluArgTrpLysSerValSerGlyValAlaAspA 240
1188 GGTAACTAAACTGGTCGAGCGATGATTCCTCTCTGTGTAGCTGATG 1237
240 SPProAsnAsnTrpLeuPheCysArgValArgLysAsnGlyValAlaAla 256
1238 ATCCGATTAATACACTACCTGTTTCCGGGTAGAAAATAGTGTCCCGCG 1287
257 ProSerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyLLePheG1 273
1288 CCATTCGCCACACGACGCTATCACTCGCGCCCTGGAAAGGATTTTGA 1337
273 uAlaThrHisArgLeuLeuTyrGlyValAlaLysAspSerGlyLinarGT 290
1338 AGCAACTCATGATGATTTACGGCGCTAAGGATGACTCTGTGCAGAGAT 1387
290 YLeuAlaTrpSerGlyHisSerAlaArgValGlyValAlaArgAspMet 306
1388 ACCTGCCCTGGTCTGACACACACTGCCGTGCGAGCCCGCGAGATATG 1437
307 AlaArgAlaGlyValSerLeuProGluLeuMetGlnAlaGlyLTrpTh 323
1438 GCCCGCGCTGGAGTTTCATACCGGAGATCATGCAAGCTGGTGGCTGAC 1487
323 rAsnValAsnIleValMetAsnTrpLLeArgAsnLeuAspSerGluThrG 340
1488 CAATTAATAATGTCATGAACATAATCCGTAACCTGTGATGTAAGAACAG 1537
340 YAlaMetValArgLeuLeuGluAspGlyAsp 350
1538 GGGCAATGCTGGCGCTGCGAAGATGGCGAT 1569

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAD09259
seq_documentation_block:
ID AAD09259 standard; DNA; 2004 BP.
XX
AC AAD09259;
XX
DT 12-SEP-2001 (first entry)
XX
DE VP22-Cre fusion DNA.
XX
KW DNA recombinase domain; protein transduction domain; PTD;
KW gene alteration; VP22-Cre fusion protein; Human immunodeficiency virus;
KW HIV; Human spumaretrovirus; HSV; ds.
XX
OS Chimeric - Human spumaretrovirus.
XX
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..2004
FT /tag= a
FT /product= "VP22-Cre fusion protein"
XX
PN WO200149837-A2.
XX
PD 12-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-EP00060.
XX
PR 07-JAN-2000; 2000EP-0100351.
XX
PR 10-NOV-2000; 2000EP-0124595.
XX
PA (ARTE-) ARTEMIS PHARM GMBH.
XX
PI Schwenk F;
XX
DR WPI; 2001-441873/47.

```

```

DR P-PSDB; AAE05266.
XX
XX using site-specific DNA recombinase domain/protein transduction domain
PT fusion proteins for inducing target gene alterations in organisms or
PT cell cultures -
XX
XX Claim 17; Page 32-35; 85pp; English.
XX
XX The present invention relates to use of fusion proteins comprising
CC a site-specific DNA recombinase domain e.g. Cre and a protein
CC transduction domain (PTD) e.g. the Human immunodeficiency virus
CC (HIV) derived TAT peptide, for preparing an agent for inducing
CC target gene alterations in a living organism or cell culture. The
CC present invention also provides a method for inducing gene
CC alterations in living organisms using the fusion proteins of the
CC invention. The present sequence is a DNA encoding VP22-Cre fusion
CC protein. The VP22 sequence is from Human spumaretrovirus (HSV).
XX
SQ Sequence 2004 BP; 419 A; 608 C; 592 G; 385 T; 0 other;

alignment_scores:
Quality: 1759.00 Length: 344
Ratio: 5.113 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.419

Alignment block:
US-09-662-128A-2 x AAD09259 ..

Align seg 1/1 to: AAD09259 from: 1 to: 2004

7 LysValSerAsnLeuThrValHisGlnAsnLeuProAlaLeuProVa 23
.....
970 GCGATGTCAAATTACTGACCGTACACCAAAATTTGCCCTCATTTACCGGT 1019

23 lAspAlaThrSerAspGluValArgLysAsnLeuMetAspMetPheArg 40
.....
1020 CGATGCACACGAGTATGAGTGTCCCAAGAACCTGATGACATGTCAAGG 1069

40 sPArgGlnAlaPheSerGluHisThrTrpLysMetLeuSerValCys 56
.....
1070 ATCCGACAGCGCTTTCTGACCATCTCCGGAATAATGCTTGTCCGTTGG 1119

57 ArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArgLysTrpPheProAl 73
.....
1120 CGTCTGCGGCGCGCATGTCTCAAGTTGAATTAACCGAATATGTTTCCCGC 1169

73 aGluProGluAspValArgAspTrpLeuLeuTrpLysGlnAlaArgGlyL 90
.....
1170 AGAACCTGAAGATGTTCCGATATATCTTCTATATCTTCAGCGCGCGGTC 1219

90 eAlaValLysThrIleGlnHisLeuGlyGlnLeuAsnMetLeuHis 106
.....
1220 TGGCAGTAAAACTATCCAGCAACATTTTGCCACGCTAAACATGCTTCAT 1269

107 ArgArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuVa 123
.....
1270 CGTCTGCTCGCGCTGCCACGACCAAGTGCACCAATGCTGTTCACTGGT 1319

123 lMetArgArgIleArgLysGluAsnValAspAlaGlyLeuArgAlaLysG 140
.....
1320 TATGCGCGGATCCGAAAAGAAACGTTGATCCGCTGACATGCAGAAAC 1369

1320 lAlaLeuAlaPheGluArgTrpAspPheAspGlnValArgSerLeuMet 156
.....
1370 AGGCTCTAGCGTTGCAACGCACTGATTTGACCAAGTTCGTTCACTCATG 1419

157 GluAsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyL 173
.....
1420 GAATAATGCGATCGCTCCAGGATATACGTATATCGCATTTCTGGGGAT 1469

173 eAlaTrpAsnThrLeuValArgIleAlaGluIleAlaArgIleArgValL 190
.....

```



```

157  GluAsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyI 173
1420  GAAATAGGCGATCGCGCAGCATATACCTATCTGGCATTTCTGGGGAT 1469
173  eAlaIYrAsnThrLeuLeuArgIleAlaGluIleAlaIArgIleArgVal 190
1470  TGCTTATTAACACCCCTGTACGTATAGCCGAATTCAGATCGAGGTTA 1519
190  yAspIleSerArgThrAspGlyGlyArgMetLeuIleHisIleGlyArg 206
1520  AAGATATCTCAGCTACTGACGGTGGGAGAACTTATCATCAATTGGGAGA 1569
207  ThrIYsThrLeuValSerThrAlaGlyValGluYsAlaLeuSerLeuG 223
1570  ACGAAMACCGCTGTAGCACCCAGCTAGAGAAAGCACTTACCCTGGG 1619
223  yValThrIYsLeuValGluArgTrpIleSerValSerGlyValAlaAsp 240
1620  GGTAACATAACCTGGTGGAGCGATGATTCCTCTGTGGTGTACTGATG 1669
240  sPProAsnAsnIYrLeuPheCysArgValArgIYsAsnGlyValAla 256
1670  ATCCCAATTAACCTACTGTTTCCGGGGTCACAAAAAATGGTGTCCGCCG 1719
257  ProSerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheG 273
1720  CCATCTGCCACCGACCACTATCACTCGCCCTCGAAGAGATTTTGA 1769
273  uAlaThrIAsArgLeuIleYrGlyAlaIYsAspAspSerGlyGlnArgT 290
1770  AGCAACTCATGATTCATTTACGGCGCTAAGATGACTCTGGTCAGAGAT 1819
290  YrIleuAlaTrpSerGlyHisSerAlaArgValGlyAlaIAspMet 306
1820  ACCTGGCGGTGTGGACACAGTCCCGTGTGGAGCCGCGCAGATATG 1869
307  AlaArgAlaGlyValSerIleProGluIleMetGlnAlaGlyIYrTrpH 323
1870  GCCCGCGGTGAGTTTCATATCCGAGATCATGCAAGCTGGTGGTGAGC 1919
323  rAsnValAsnIleValMetAsnTYrIleArgAsnLeuAspSerGlyUThrG 340
1920  CAATGTAAATTTGTCAATGAACTATATCCGTAACTGGATGAGCAACAG 1969
340  IYAlaMetValArgLeuLeuGluAspGlyAsp 350
1970  GGCAATGTCGCGCTCTGTGAAAGATGGCGAT 2001

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAD09266
seq_documentation_block:
ID  AAD09266 standard: DNA: 4488 BP.
XX
AC  AAD09266;
XX
DT  12-SEP-2001 (first entry)
XX
DE  PT7-VPCS vector DNA related to the invention.
XX
KW  DNA recombinase domain; protein transduction domain; PTD;
KW  gene alteration; fusion protein; Human immunodeficiency virus;
KW  HIV; PT7-VPCS vector; ds.
XX
OS  Unidentified.
XX
PN  WO200149832-A2.
XX
PD  12-JUL-2001.
XX
PF  05-JAN-2001; 2001WO-EP00060.
XX
PR  07-JAN-2000; 2000EP-0100351.

```

```

PR  10-NOV-2000; 2000EP-0124595.
XX
XX  (ARTE-) ARTEMIS PHARM GMBH.
XX
PI  Schwenk F.
XX
DR  WPI: 2001-441873/47.
XX
PT  Using site-specific DNA recombinase domain/protein transduction domain
PT  fusion proteins for inducing target gene alterations in organisms or
XX  cell cultures -
XX
PS  Example; Page 51-52; 85pp; English.
XX
CC  The present invention relates to use of fusion proteins comprising
CC  a site-specific DNA recombinase domain e.g. Cre and a protein
CC  transduction domain (PTD) e.g. the Human immunodeficiency virus
CC  (HIV) derived TAR peptide, for preparing an agent for inducing
CC  target gene alterations in a living organism or cell culture. The
CC  present invention also provides a method for inducing gene
CC  alterations in living organisms using the fusion proteins of the
CC  invention. The present sequence is a PT7-VPCS vector DNA related to
XX  the invention.
XX
SQ  Sequence 4488 BP; 1038 A; 1239 C; 1190 G; 1021 T; 0 other:

alignment_scores:
      Quality: 1759.00      Length: 344
      Ratio: 5.113          Gaps: 0
Percent Similarity: 100.000  Percent Identity: 99.419

alignment_block:
US-09-662-128A-2 x AAD09266 ..

Align seg 1/1 to: AAD09266 from: 1 to: 4488

7  LysValSerAsnLeuLeuThrValIleHisGlnAsnLeuProAlaLeuProAla 23
   ::::|||||
3369  CGCATGTCCAAATTTACTGACCGTACACCAAAATTTGCTGCTCATTTACCGGT 3418

23  lAspAlaThrSerAspGluValArgIYsAsnLeuMetAspMetPheArgA 40
   |||||
3419  CGATGCAACAGATGATGAGTGTGCAAGAACCTGATGACATGTTGACGG 3468

40  sPArgGlnAlaPheSerGlnHisThrTYrIYsMetLeuLeuSerValYs 56
   |||||
3469  ATCCCGAGGGGTTTCTGACCATACCTGGAAATGCTTGTCCGCTTTC 3518

57  ArgSerTrpAlaAlaTrpCysLYsLeuAsnAsnArgLYsTrpPheProAl 73
   |||||
3519  CGGTGCGTGGCGGCATGTGTCAGTGTGATTAACCGAAATGTTTCCCGC 3568

73  aGluProGluAspValArgAspTYrLeuLeuTYrLeuGlnAlaArgGlyL 90
   |||||
3569  AGAACCTGAAGATGTTCCGATATATCTTATATGTTTCAGCGCGCGGTC 3618

90  eValAlaValYsThrIleGlnGlnHisLeuGlyGlnLeuAsnMetLeuHis 106
   |||||
3619  TGGCAGTAAATACTATCCAGCAACATTTGGGCCACTTAAACATGCTTCAT 3668

107  ArgArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuVal 123
   |||||
3669  CGTGGTCCGGGCTGCCACGACCAAGTGCACACATGCTGTTTCACTGCT 3718

123  lMetArgArgIleArgLYsGluAsnValAspAlaGlyGluArgAlaYsG 140
   |||||
3719  TATGCGCGGATCCGAAAGAAAGACGTTGATGCCGTGAACGTCCAAAC 3768

140  lAlaLeuAlaPheGluArgThrAspPheAspGlnValArgSerLeuMet 156
   |||||
3769  AGGCTCTACGTTTCGACGCACTGATTTTCGACCAAGTTCGTTCACTCATG 3818

```

```

157  GUAASerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGly11 173
      |||||||
3819  GAAATATGCGATCGCTGCCAGGATATACGTAATCTGGCATTTCTGGGAT 3868
      |||||||
173  eAlaTyrAsnThrLeuAlaArgIleAlaGlyIleArgVal1 190
      |||||||
3869  TGGTTATTAACCCCTGTAGCATATAGCCGAAATTCGACGATAGGTTA 3918
      |||||||
190  yAspIleSerArgThrAspGlyGlyArgMetLeuIleHisIleGlyArg 206
      |||||||
3919  AAGATATCTCAGTACTGACGCTGGAGAAATGTTAATCCATATTTGGCAGA 3968
      |||||||
207  ThrIlyThrLeuValSerThrAlaGlyAluLysAlaLeuSerLeuG1 223
      |||||||
3969  ACGAAACCGTGTAGACCGCAGCGTGTAGAGAGGACCTAGCCCTGGG 4018
      |||||||
223  yAlaThrIleuValGluArgTrpIleSerValSerGlyValAlaAspA 240
      |||||||
4019  GGTAACTAACCTGGTCGAGCGATGATTTCCGTTCTGTGGTGTAGCTGATG 4068
      |||||||
240  sPProAsnAsnTyrLeuPheCysArgValAlaArgLysAsnGlyValAla 256
      |||||||
4069  ATCCGAATTAACCTGTTTGGCCGGGTACGAAATAATGGTGTGGCCGC 4118
      |||||||
257  ProSerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheG1 273
      |||||||
4119  CCATCTGCCACACGACGACTATCAACTGCGCCCTGGAGAGGATTTTGA 4168
      |||||||
273  uAlaThrHisArgLeuIleTyrGlyAlaAlaLysAspSerGlyGlnArgT 290
      |||||||
4169  ACCAACTCATGATGATTAGGGGGCTAAGAGTACTCTGTGCAGAGAT 4218
      |||||||
290  yLeuAlaThrSerGlyHisSerAlaArgValGlyAlaAlaArgAspMet 306
      |||||||
4219  ACCGTGCCCTGCTGTGACACATGCGCGTGTGGAGCCGCGAGATATG 4268
      |||||||
307  AlaArgAlaGlyValSerIleProGluIleMetGlnAlaGlyIleTrpTh 323
      |||||||
4269  GCCCGCGCTGGAGTTTCATACCGAGATACGCAAGCTGGTGGCTGAC 4318
      |||||||
323  rAsnValAsnIleValMetAsnTyrIleArgAsnLeuAspSerGluTrpG 340
      |||||||
4319  CAATGTAAATATTTGTCATATATCCGTAAACCTGGATGATGTAACAG 4368
      |||||||
340  lValIleMetValArgLeuLeuGluAspGlyAsp 350
      |||||||
4369  GGGCAATGCTGCCCTGCTGGAAGATGGCGAT 4400
      |||||||
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAc62541
seq_documentation_block:
ID   AAC62541 standard; DNA; 1172 BP.
XX
AC   AAC62541;
XX
DT   07-FEB-2001 (first entry)
XX
DE   Cre wild-type coding sequence.
XX
KW   Cre variant recognition site; lox site; recombinase;
KW   variant recombination site; hybrid crop production; seedless crop;
KW   phage packaging; cloning; ds.
XX
OS   Unidentified.
XX
PN   WO200060091-A2.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US09154.
XX
PR   06-APR-1999; 99US-0127977.
XX

```

```

PA   (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX
PI   Sauer BL, Rufer AW;
XX
DR   WPI; 2000-665010/64.
XX
PT   Identifying variant recombinases mediating recombination at variant
PT   sites (VRS) by contacting a mutant recombinase, a first and second VRS
PT   having a reporter gene, and a second nucleic acid having 2 VRS and a
PT   reporter gene -
XX
PS   Example 1; Page 98-101; 144pp; English.
XX
CC   The present invention relates to the identification of recombinase
CC   variants which have an altered specificity. They are tested using
CC   constructs containing variant recognition sites, which are not recognised
CC   by non-mutant recombinase but undergo recombination in the presence of a
CC   variant enzyme. Variant recombinases are useful in the production of
CC   genetically modified crop plants, particularly seedless varieties, and in
CC   phage packaging, which has uses in cloning.
XX
SQ   Sequence 1172 BP; 299 A; 263 C; 323 G; 287 T; 0 other;

```

```

alignment_scores:
      Quality: 1758.00      Length: 344
      Ratio: 5.110          Gaps: 0
Percent Similarity: 100.000      Percent Identity: 99.419

```

```

alignment_block:
US-09-662-128a-2 x AAC62541 ..

```

```

Align seg 1/1 to: AAC62541 from: 1 to: 1172

```

```

7  LysValSerAsnLeuThrValHisGlnAsnLeuProAlaLeuProVa 23
      |||||||
53  GAAATGTCATTTCTGACCGTACACCAAAATTTGCTGCATTAACCGGT 102
      |||||||
23  lAspAlaThrSerAspGluValArgLysAsnLeuMetAspMetPheArgA 40
      |||||||
103  CGATCGAACAGTGTAGTGGTCCGCAAGACCTGATGACATGTGCGAGG 152
      |||||||
40  sPArgGlnAlaPheSerGluHisThrTrpLysMetLeuLeuSerValCys 56
      |||||||
153  ATGCGCAGCGCTTTCTGACGATTAACCTGGAATAAGCTCTGCTCGTTGC 202
      |||||||
57  ArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArgLysTrpPheProAl 73
      |||||||
203  CGGTGCTGGCGCGCATGTGCAAGTTGAATACCGGAATGTTCCCGC 252
      |||||||
73  eGluProGluAspValArgAspTyrLeuLeuTyrLeuGlnAlaArgGlyL 90
      |||||||
253  AGAACCTGAAGATGTCGCGATTAATCTTATATTTAGGCGCGCGGTC 302
      |||||||
90  eAlaValIlyThrIleGlnGlnHisLeuGlyGlnLeuAsnMetLeuHis 106
      |||||||
303  TGGCAGTAAAACTATCCAGCAACTTTGGCCACGCTAAACATGCTTCAT 352
      |||||||
107  ArgArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuVa 123
      |||||||
353  CGTCCGTCCGGCTGCCACGACCAAGTACAGCAATGCTGTTCACGTGCT 402
      |||||||
123  lMetArgArgIleArgGlyGluAsnValAspAlaGlyGluArgAlaLysG 140
      |||||||
403  TATGCGCGGATCCGAAGAAAGAAACGTTATGCCGTTGACGTGCAAAAC 452
      |||||||
140  lAlaLeuAlaPheGluArgTrpAspPheAspGluValArgSerLeuMet 156
      |||||||
453  AGGCTTAGCGTTCCGAACGACATGATTCGACAGGTTGTTCACTCATG 502
      |||||||
157  GUAASerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGly11 173
      |||||||
503  GAAATATGCGATCGCTGCCAGGATATACGTAATCTGGCATTTCTGGGAT 552
      |||||||

```

173 eAlaTyrAsnThrLeuLeuArgIleAlaIleAlaArgIleArgVal 190
|||||
553 TGCTATTAACACCCCTGTACGTATAGCCGAATTCAGAGATCAGGTTA 602
190 yAspIleSerArgThrAspGlyGlyArgMetLeuIleHisIleGlyArg 206
|||||
603 AAGATATCTCAGCTACTGACGTGGAGGAGATTAAATCCATATTGGCAGA 652
207 ThrLysThrLeuValSerThrAlaGlyValGluLysAlaLeuSerLeu 223
|||||
653 ACGAAACGCTGGTGTAGCACCCGAGCTGTAGAGAGGCACTTAGCCTGG 702
223 yValThrLysLeuValGluArgTrpIleSerValSerGlyValAlaAsp 240
|||||
703 GGTAACTAACTGTCGACGAGATGATTCGCTCTGGGTAGCTGATG 752
240 sProAsnAsnTyrLeuPheCysArgValArgLysAsnGlyValAlaAla 256
|||||
753 ATCCGAATTAACCTACTGCTGTTTGCCGGGTCAAAAAAATGTGTGCCCG 802
257 ProSerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheG 273
|||||
803 CCATCTGCCACCAAGCAGCTATCAACTCGCCCTGGAGAGGATTTTGA 852
273 uAlaThrHisArgLeuIleTyrGlyAlaLysAspAspSerGlyGlnArg 290
|||||
853 AGCAACTCATGATGATTACGGCGCTAAGATGATCTCGTCAGAGAT 902
290 yrLeuAlaTrpSerGlyHisSerAlaArgValGlyAlaAlaArgAspMet 306
|||||
903 ACCTGGCCGTGCTGTGACACAGTGCCTGTGTGGAGCCGCGGAGATATG 952
307 AlaArgAlaGlyValSerIleProGluIleMetGlnAlaGlyTyrTrp 323
|||||
953 GCCCGCGGTGAGTTCAATACCGAGATCATGCAAGCTGTGGCTGAGAC 1002
323 rAsnValAsnIleValMetAsnTyrIleArgAsnLeuAspSerGluThr 340
|||||
1003 CAATGTAAATATGTGCATGAACCTATATCCGTAAACCTGATAGTAAACAG 1052
340 yAlaMetValArgLeuLeuGluAspGlyAsp 350
|||||
1053 GGGCAATGTGTGGCTGTGGAAGATGGCGAT 1084

OM of: US-09-662-128a-2 to: Issued_Patents_NA:* out_format : pfs
Date: May 10, 2002 4:10 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=firmet_p2n.model -DEV=xip
-O=/cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-457-342-1 + 86.00 126.02 24.86
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=tbl -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsun62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pcp
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPR=pfis
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US09662128.ecgn1_1.85 -NCPV=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-662-128a-2
Query length: 350
Database: Issued_Patents_NA:*
Database sequences: 38353
Database length: 122816752
Search time (sec): 71.710000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	5.5e-197	1032
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-193-503B-2		1757.00	3674.50	1		1.3e-196	1740
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-864-224-10		1757.00	3668.04	1		2.0e-196	2346
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-193-503B-5		1757.00	3668.04	1		4.6e-196	2346
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-193-503B-4		1754.00	3658.01	1		4.6e-196	2346
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-193-503B-8		1754.00	3658.01	1		4.6e-196	2346
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-735-609-11		671.00	1388.52	1		1.2e-69	699
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-315-372-11		671.00	1388.52	1		1.2e-69	699
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-244-752-11		671.00	1388.52	1		1.2e-69	699
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-245-497-11		671.00	1388.52	1		1.2e-69	699
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-105-390-51		95.50	135.78	1		7.11	16075
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-096-867-11		95.50	135.78	1		7.11	16075
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-814-052-12		92.00	159.62	0.3340		0.3340	1288
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-743-637B-171		92.00	159.62	0.3340		0.3340	1288
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-318-448-2		91.50	137.41	5.77		7122	1288
/cgn2_6/prodata/2/lna/5A_COMB.seq:US-09-103-840A-1		91.50	134.61	8.26		8931	1
/cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-764-233A-1		91.50	113.45	124.57		49377	1
/cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-110-840A-2		90.00	151.62	0.9324		4403765	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-178-019-1		89.00	142.62	2.96		3619	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-105-390-2		89.00	149.12	1.97		11958	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-134-246-8		89.00	126.78	22.55		4403765	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-103-840A-2		89.00	52.62	2.5e+05		11958	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-457-342-1		88.50	125.09	28.01		11561	1
/cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-614-770A-1		88.50	107.15	29.54		49272	1
/cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-814-052-13		88.00	148.35	1.42		1620	1
/cgn2_6/prodata/2/lna/5A_COMB.seq:US-07-718-575-13		88.00	139.38	4.48		3344	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-481-206-13		88.00	139.38	4.48		3344	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-269A-13		87.50	49.43	3.4e+05		4411529	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-103-840A-1		86.00	127.13	5.98		2852	1
/cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-258-261B-1		86.00	126.02	24.86		7001	1
/cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-456-837-1		86.00	126.02	24.86		7001	1

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-193-503B-2
seq_documentation_block:
; Sequence 2, Application US/09193503B
; Patent No. 6262341
; GENERAL INFORMATION:
; APPLICANT: Basczinski, Christopher L.
; APPLICANT: Lysnik, Leszek A.
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Guan, Xueni
; APPLICANT: Rao, Guru
; APPLICANT: Taglianti, Laura A.
; TITLE OF INVENTION: A No. 6262341el Method For The Integration Of Foreign DNA Int
; FILE REFERENCE: 5718-66 (amended listing)
; CURRENT APPLICATION NUMBER: US/09/193,503B
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/099,435
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: 60/056,627
; PRIOR FILING DATE: 1997-11-18
; PRIOR APPLICATION NUMBER: 60/065,613
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
; OTHER INFORMATION: sequence encoding Cre protein from Bacteriophage
US-09-193-503B-2

alignment_scores:
Quality: 1757.00 Length: 343
Ratio: 5.122 Gaps: 0
Percent Similarity: 100.00 Percent Identity: 99.708

alignment_block:
US-09-662-128a-2 x US-09-193-503B-2 ..

Align seg 1/1 to: US-09-193-503B-2 from: 1 to: 1032

8 ValSerAsnLeuThrValHisGlnAsnLeuProAlaLeuProValAs 24
1 ATGTCAACCTGCTCAGGTCACGACGACCTTCGCGCTTCAGTGA 50
24 palatirSerAspGluValArgGlyAsnLeuMetAspMetPheArgAsp 41
51 CGGACGCTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 100
41 rglAlaPheSerGluHisThrTrpGlyMetLeuLeuSerValGlyArg 57
101 GGCACACGCTTACGCGACGACCTGAGAGTCTGCTCTCCGTCGCCG 150
58 SerTrpAlaAlaTrpGlyLeuAsnGlnAsnArgGlyTrpPheProAlaG 74
151 TCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 200
74 ubProGluAspValArgAspTrpLeuLeuTrpLeuGlnAlaArgGlyLeu 91
201 GCCCGGACGCTGAGGATTTACCTTCTGACTGCAACCTGCGGCGCTG 250
91 laValLysThrIleGlnHisLeuGlyGlnLeuAsnMetLeuHisArg 107


```

|||||
964 AAAACATCCAGACATTTGGGCGACGTAACATGCTTCATCGTGGTC 1013
109 TGLYLEPRARIGROSERASPSERSNALVALSERLEUVALMETARGA 126
1014 CGGGCTGCCAGCACAGTAGACAGCAATGCTGTTCCACGTTAAVGGCGC 1063
126 TGLLEARGLSGLUASNVALASPALAGLYLUARGALATYSGINLALAU 142
1064 GGATCCGAAAGAAACGTTAGCCGGGGAACGTCGCAAAACAGCTCTTA 1113
143 ALAPHGLUARGTHRASPHEASPGINVALARGSERLEUMETGLUASNE 159
1114 GCGTTCGAACGACACTGATTTCCAGCCAGGTCCTCATCATGGAATAATG 1163
159 RASPARCYSGINASPILEARGASNLEUALAPHELEUGLYLEALATYRA 176
1164 CGATCGCTCCAGGATATACGTAATCTGGCATTTCTGGGAGATTGCTTATA 1213
176 SRRHRLLEUARGYLEALAGLUILEALARGILEARGVALLYSASPILE 192
1214 ACACCTGTTAGCTATAGCCGAAATTCAGAGATCAGGCTTAAGATATC 1263
193 SERARGTHRASPGLYGLYARGMETLEULIENHISIEGLYARGTHRLYSTH 209
1264 TCACGACTAGCAGCGTGGGAGAAATGTTAAATCCATATTGGCAGAACGAAAC 1313
209 RLEUVALSERHTRALAGLYVALGLULYSALALEUSERLEUGLYVALTHRL 226
1314 GCTGGTTACACCGCAGGTGTAGAGAACGACTAGCTGGGGGTAACATA 1363
226 YLEUVALGLUARGTRPTLESERVALSERGLYVALALASPPROASN 242
1364 AACCTGTCAGCGAGTAGATTCCGTCCTGTTAGCTGATGATCCCAAT 1413
243 ASANTYRLEUPHECYARGVALARGLYSASNGLYVALALIAALPROSERAL 259
1414 AACCTACCTGTTTGGCGGGTCAGAAATAATGCTGTGCGCGGCACACTGC 1463
259 ARHRSERGLNEUSERHTRARGALALEUGLYLLEPHEGLUALATHRH 276
1464 CACGACGCAAGTATCACTCCGCCCTGGAGGAGGATTTTGAACCACTC 1513
276 ISARGLEULIETRYGLYALALYASPSASPSERGLYGINARGTRYLEUALA 292
1514 ATCGATTGATTTACGGCGCTAAGATGACCTGCTCAGAGATACCTGGCC 1563
293 TRPSERGLYHISERVALARGVALGLYVALALARGASPMETALAARGAL 309
1564 TGGTCTGGACACAGTCCCGCTGTGGAGCCGCGGAGATATGGCCGCCG 1613
309 AGLYVALSERILEPROGLIULIEMETGINALAGLYLTRYPTHRASNVALA 326
1614 TGGAGTTTCATACCGAGATACGCAAGCTGTGGCTGGACCAACTGTAA 1663
326 SNLIEVALMETASNTRYLLEARGSNLEUASPSERGLUTHRGLYALAMET 342
1664 ATATTGTCATGAACATATACCGTAACCTGATAGTGAACAGGGCNAATG 1713
343 VALARGLEULEUGLUASPGLYASP 350
1714 GTGCGCTGCTGGAGATGGCGAT 1737
seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-09-193-503B-5
seq_documentation_block:
; Sequence 5, Application US/09193503B
; Patent No. 6262341
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Christopher L.
; APPLICANT: Lyznik, Leszek A.
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Guan, Xueni

```

```

; APPLICANT: Rao, Guru
; APPLICANT: Tagliani, Laura A.
; TITLE OF INVENTION: A No. 6262341e1 Method For The Integration Of Foreign DNA Int
; TITLE OF INVENTION: Eukaryotic Genomes
; FILE REFERENCE: 5718-66 (amended listing)
; CURRENT APPLICATION NUMBER: US/09/193,503B
; CURRENT FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/099,435
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: 60/056,627
; PRIOR FILING DATE: 1997-11-18
; PRIOR APPLICATION NUMBER: 60/065,613
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO: 5
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: sequence
; OTHER INFORMATION: encoding mCre:FlpM, Cre from Bacteriophage P1 and
; OTHER INFORMATION: FLP from Saccharomyces, both maize preferred
; NAME/KEY: CDS
; LOCATION: (1)..(2346)
US-09-193-503B-5

```

alignment_scores:

```

Quality: 1757.00 Length: 343
Ratio: 5.122 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.708

```

alignment_block:

```

US-09-662-128a-2 x US-09-193-503B-5 ..

```

```

Align seg 1/1 to: US-09-193-503B-5 from: 1 to: 2346

```

```

8 VALSERASNLEUETHRVALLHISGLASNLEUPROALALEUPROVALAS 24
1 ATGTCCAACTGCTGCACGGTTCACAGAACCTTCGCGCTTCACAGGA 50
24 PALATHSERASPGILUVALARGLYSASNLEUMETASPMETPHATHASPA 41
51 CGCGACGTCGATGAGTCAGCAAGAACCTCATGAGACATGTCGCGACA 100
41 RGLNALAPHSESERGLUHISTHTRPTLYSMETLEUSERVALICYARG 57
101 GCCAAGCGTTCAGCGACACACCTGGAAGATGCTGCTCCGCTCGCCG 150
58 SERTPALAALATPRCYLSYLSLEUASNARGLYSTRPHEPROALAGI 74
151 TCTCGGGCTGCATGCTGCAGCTGAACACAGAGAGTGGTCCCGCTGA 200
74 UPROGLUASPYVALARGAPTRYLEUETHRYLEUGINALARGGLYLEUA 91
201 GCCCGAGGAGCGTAGGAGATTACCTTCTGTACTGTCGAACCTCGCGGCTG 250
91 LVALLYSTRHLLIEGLINGHISLEUGLYSLNLEUASNMETLEUHIARG 107
251 CAGTGAAGACCATCCAGCAACACCTTGGACACTGAACATGCTTCACAG 300
108 ARGSERGLYLEUPROARGPROSERASPSERASNALVALSERLEUVALME 124
301 CGCTCCGGCTCCCGCGCCCGCCAGCAGACTGAAACGCCGTACCTGTCA 350
124 TARGARGILEARGLYSGLUASNVALASPALAGLYLUARGALATYSGINA 141
351 GCGCGCGCATCAGGAAGAAACCTCGATGCTCCGCGAAAGCGCAAGG 400
141 LALUVALAPHEGLUARGTHRASPPHEASPGINVALARGSERLEUMETGLU 157

```

```

401 CCTCGCGTTCGAGAGACGATTTCACACAGGTCGCCACCTGATGGAG 450
158 AsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyIleAl 174
|||||
451 AACACGACGAGGTGCGACGACATTAGGACCTGGCGTTCTCGGAATTGC 500
174 aTyrAsnThrLeuLeuArgIleAlaGluIleAlaArgIleArgValIysA 191
|||||
501 ATACAAACAGCGTCTCCAGAGATCGGGAAATTGCCGCAATTCGGCTGAAG 550
191 spIleSerArgThrAspGlyIleArgMetLeuIleHisIleGlyArgThr 207
|||||
551 ACATTAGCGCGACCGACGCGCGGACGAGATGCTTATCCACATTGGCAGGACC 600
208 LysThrLeuValSerThrAlaGlyValGluIysAlaLeuSerLeuGlyVa 224
|||||
601 AAGAGCGCTGTTTCCACCGCAGAGCGCTCGAAAAGGCCCTCAGCCTCGGAGT 650
224 IThrIysLeuValGluArgTrpIleSerValSerGlyValAlaAspAsp 241
|||||
651 GACCAAGCTCGTGAACGCTGATCTCGTGTCGGCGCTCGCGACGACGCC 700
241 rOAsnAsnIleuPheCysArgValArgLysAsnGlyValAlaAlaPro 257
|||||
701 CAACACACTACTTCTTGCGCGCTCCGCAAGAACGGGGTGGTGGCCCT 750
258 SerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheGluAl 274
|||||
751 AGGGCCACAGCCCACTCAGCCAGAGGCGCTGGAAAGTATTTCGAGGC 800
274 aThrHisArgLeuIleTyrGlyAlaIysAspAspSerGlyIleArgTyrL 291
|||||
801 CACCCACCGCGTATCTACGCGCGGCAAGATGACAGCGGTCAACGCTACC 850
291 euAlaTrpSerGlyHisSerAlaArgValGlyAlaAlaArgAspMetAla 307
|||||
851 TCCATAGTTCGGGCGACCTCCGCGCGGTGGAGTCTAGGACATGGCGC 900
308 ArgAlaGlyValSerIleProGluIleMetGlnAlaGlyIleTrpThrAs 324
|||||
901 CGGCGCGGTGTTCCATCCCGAAATCATGCAAGCGGGGTGATGACGAA 950
324 nValAsnIleValMetAsnTyrIleArgAsnLeuAspSerGluThrGlyA 341
|||||
951 CGTGAACATGTGATGACTGACATTGCAACCTTGACAGCAGAGCGGGCG 1000
341 IametValArgLeuLeuGluAspGlyAsp 350
|||||
1001 CAATGGTTCGCTCTCGAGAGATGGCGAT 1029
seq_name: /cgn2_6/ptodata/2/ina/6B-COMB.seq:US-09-193-503B-4
seq_documentation_block:
; Sequence 4, Application US/09193503B
; Patent No. 6262341
; GENERAL INFORMATION:
; APPLICANT: Baszczyński, Christopher L.
; APPLICANT: Lyszniak, Leszek A.
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Guan, Xueni
; APPLICANT: Rao, Guru
; APPLICANT: Taglianti, Laura A.
; TITLE OF INVENTION: A No. 6262341el Method for The Integration of Foreign DNA Into
; FILE REFERENCE: 5718-66 (amended listing)
; CURRENT APPLICATION NUMBER: US/09/193,503B
; CURRENT FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/099,435
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: 60/056,627
; PRIOR FILING DATE: 1997-11-18
; PRIOR APPLICATION NUMBER: 60/065,613
; PRIOR FILING DATE: 1997-11-18

```

```

; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: sequence encoding a Cre:FLPm polypeptide, Cre
; OTHER INFORMATION: from Bacteriophage p1 and FLP (maize preferred
; OTHER INFORMATION: codons) from Saccharomyces
; US-09-193-503B-4

```

```

alignment_scores:
    Quality: 1754.00      Length: 343
              Ratio: 5.114      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 99.417

```

alignment_block:

US-09-662-128A-2 x US-09-193-503B-4

Align seg 1/1 to: US-09-193-503B-4 from: 1 to: 2346

```

8 ValSerAsnLeuLeuThrValHisGlnAsnLeuProAlaLeuProValAs 24
|||||
1 ATGGCCCAATTACTACCCGCTACACCAAAATTGGCTCATTCACCGGTGCA 50
24 PalArThrSerAspGluValArgLysAsnLeuMetAspMetPheArgAspA 41
|||||
51 TGCACAGAGTATGAGTTCGCAAGAACCTGATGACATGTTACGGGATC 100
41 rGlnAlaPheSerGluHisThrTrpLysMetLeuLeuSerValCysArg 57
|||||
101 GCCAGGCGTCTTCTGAGCATACCTGGAATAATGCTTGCTGCTTGCCGG 150
58 SerTrpAlaIleTrpCysLysLeuAsnAsnArgLysTrpPheProAlaG 74
|||||
151 TCGTGGCGCGCATGTGTCAGAGTTGAATACCGGAATGTTCCCGCAGA 200
74 uProGluAspValArgAspTrpIleuLeuTyrLeuGlnAlaArgGlyLeuA 91
|||||
201 ACCTGAAGATGTTGGCGATTATCTTATATCTTACGCGCGCGGTCTGG 250
91 IValIysThrIleGlnGlnHisLeuGlyGlnLeuAsnMetLeuHisArg 107
|||||
251 CAGTAAAACTATCCAGCAACATTGGGCCAGCTTAAATGCTTCAATCGT 300
108 ArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuValMe 124
|||||
301 CGGTCCGGCGCTGCCAGACCAAGTGACACAAATGCTGTTCACTGTTAT 350
124 ValArgGluIleArgLysGluAsnValAspAlaGlyIleuArgAlaLysGlnA 141
|||||
351 GCGGCGGATCCGAAAGAAACGTTGATGCCGTGAACGTCGAAACACGAG 400
141 ILeuAlaPheGluArgTrpAspPheAspGluValArgSerLeuMetGlu 157
|||||
401 CTCTAGAGTTCCGAACGACATGATTCGACAGAGTTCGTTCACTCATGAA 450
158 AsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyIleAl 174
|||||
451 AATAGCATGCGTCCGACGATATACGTAATCTGGCATTTCTGGGATATGC 500
174 aTyrAsnThrLeuLeuArgIleAlaGluIleAlaArgIleArgValIysA 191
|||||
501 TTATTAACACCTGTTACGATATACCGAAATGGCAGAGATCAGGGTTAAG 550
191 spIleSerArgThrAspGlyIleArgMetLeuIleHisIleGlyArgThr 207
|||||
551 ATATCTCAGTACGACGAGTGGGAGATGTTATTCATATGTCAGAGAACG 600
208 LysThrLeuValSerThrAlaGlyValGluIysAlaLeuSerLeuGlyVa 224

```

```

601 AAAACGCTGTTAGCACCGAGGTGAGAGAAAGCACTTACCTGGGGGT 650
224 lThrLysLeuValGluArgTrpLleSerValSerGlyValAlaAspAsp 241
651 AACCTAACTGTCGACGAGATGATTCCTCTGCTGTAGCTGATGATC 700
241 roAsnAsnTrpLeuPheCysArgValArgLysAsnGlyValAlaAlaPro 257
701 CGAATAACTGCTGTTTGGCGGGGTACGAAAAATGGTGTTCGCCGCCA 750
258 SerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheGluAl 274
751 TGTGCCACACGACGCTATCACTCCGCCCTGGAAGGATTTTGAAGC 800
274 aThrHisArgLeuIleTyrGlyAlaLysAspSerGlyGlnArgTyrL 291
801 AACTCATGATGATTTACGGCGCTAAGATGACTCTGTCTGACAGATACC 850
291 euAlaIrrSerGlyHisSerAlaArgValGlyAlaAlaArgAspMetAla 307
851 TGGCGTGGTCTGGACACAGTCCGCTGCGAGCGCGCGAGATATGGCC 900
308 ArgAlaGlyValSerLleProGluLleMetGlnAlaGlyGlyTrpThrAs 324
901 CGGGCGGAGTTTCAATACCGAGATCATGCAAGCTGGCTGGACCAA 950
324 nValAsnIleValMetAsnTyrLleArgAsnLeuAspSerGlyThrGly 341
951 TGTAAATATGTCATGAACTATATCCGTAACCTGGATGTGAAACACAGGG 1000
341 lMetValArgLeuLeuGluAspGlyAsp 350
1001 CAATGCTGCCCTGCTGGAGAGATGGCGAT 1029

seq_name: /cgn2_6/prodata/2/lna/bb_comb.seq:us-09-193-503B-7

seq_documentation_block:
? Sequence 7, Application US/09193503B
? Patent No. 6262341
? GENERAL INFORMATION:
? APPLICANT: Baszczynski, Christopher L.
? APPLICANT: Lyznik, Leszek A.
? APPLICANT: Gordon-Kamm, William J.
? APPLICANT: Guan, Xueni
? APPLICANT: Rao, Guru
? TITLE OF INVENTION: A No. 6262341el Method For The Integration Of Foreign DNA Into
? FILE REFERENCE: 5718-66 (amended listing)
? CURRENT APPLICATION NUMBER: US/09/193,503B
? PRIOR APPLICATION NUMBER: 60/099,435
? PRIOR FILING DATE: 1998-09-08
? PRIOR APPLICATION NUMBER: 60/056,627
? PRIOR FILING DATE: 1997-11-18
? PRIOR APPLICATION NUMBER: 60/065,613
? PRIOR FILING DATE: 1997-11-18
? NUMBER OF SEQ ID NOS: 11
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 7
? LENGTH: 2346
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: sequence
? OTHER INFORMATION: encoding a Cre:FLP polypeptide, Cre from
? OTHER INFORMATION: Bacteriophage P1 and FLP from Saccharomyces
US-09-193-503B-7

```

alignment_scores: Quality: 1754.00 Length: 343

```

Ratio: 5.114                      Gaps: 0
Percent Similarity: 100.000                      Percent Identity: 99.417
alignment_block:
US-09-662-128a-2 x US-09-193-503B-7
Align seg 1/1 to: US-09-193-503B-7 from: 1 to: 2346

8 ValSerAsnLeuThrValHisGlnAsnLeuProAlaLeuProValAs 24
1 ATGGCCAATTTACTGACCGCTACACCAAAATTTGGCTGCATTCACCGTGA 50
24 PalatThrSerAspGluValArgLysAsnLeuMetAspMetPheArgAspA 41
51 TGCACAGAGATGAGAGTTCGCAAGACCTGATGACATGTCACAGGATC 100
41 rglInAlaPheSerGluHisThrTrpLysMetLeuLeuSerValCysArg 57
101 GCCAGGCGTTTCTGAGCATACCTGGAATAATGCTTCTGCTCCGTTGCCGG 150
58 SerTrpAlaAlaTrpCysLysLeuAsnAsnArgLysTrpPheProAlaG 74
151 TGTGGCGCGCATGGTGCAGTTGAATACCGGAATGCTTCCCGCAGA 200
74 uProGluAspValArgAspTyrLeuLeuTyrLeuGlnAlaArgLysLeuA 91
201 ACCTGAAGATGTCGCGATTAATCTTCTATCTTCAGCGCGCGGTCTGG 250
91 lValLysThrIleGlnGlnHisLeuGlyGlnLeuAsnMetLeuHisArg 107
251 CAGTAAACATATCCAGACACATTTGGCCAGCTTAACATGCTTCACATCG 300
108 ArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuValMe 124
301 CGTCCGGCGCTGCCACGACCAAGTGACAGACAAATCTGTTACAGGTAT 350
124 tArgArgIleArgLysGluAsnValAspAlaGlyGluArgAlaLysGlnA 141
351 GCGGGCGATCCGAAAGAAACGTTGATGCCGGGAACGTGCAAAACAGC 400
141 lAlaValAlaPheGluArgThrAspPheAspGlnValArgSerLeuMetGlu 157
401 CTCTAGCGTTGCAACGACGATGATTGACACAGGTGCTTACTCATCAGGA 450
158 AsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyIleAl 174
451 AATAGCGATGCGTGCAGATATACGTAATCTGCAATTTCTGGGCGATTGC 500
174 aTyrAsnThrLeuLeuValArgIleAlaGluIleAlaArgIleArgValLysA 191
501 TTATAACACCGCTGTAGCTATACCCGAAATTTGCCAAGATCAGGGTTAAG 550
191 spLleSerArgThrAspGlyGlyArgMetLeuIleHisIleGlyArgThr 207
551 ATATCTCAGCTACTGACGGTGGGAGAAATGTTATCATATTGGCAGAACG 600
208 LysThrLeuValSerThrAlaGlyValGluLysAlaLeuSerLeuGlyVal 224
601 AAAACGCTGTTAGCACCGAGGTGAGAGAAAGCACTTACCTGGGGGT 650
224 lThrLysLeuValGluArgTrpLleSerValSerGlyValAlaAspAsp 241
651 AACCTAACTGTCGACGAGATGATTCCTCTGCTGTAGCTGATGATC 700
241 roAsnAsnTrpLeuPheCysArgValArgLysAsnGlyValAlaAlaPro 257
701 CGAATAACTGCTGTTTGGCGGGGTACGAAAAATGGTGTTCGCCGCCA 750
258 SerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheGluAl 274
751 TGTGCCACACGACGCTATCACTCCGCCCTGGAAGGATTTTGAAGC 800
274 aThrHisArgLeuIleTyrGlyAlaLysAspSerGlyGlnArgTyrL 291

```

```

|||||
801 AACTCATCGATTGATTACGGCGCTAAGATGACTCTGCTGCAGAGATACC 850
291 euA1arPserGlyHisSerAlaArgValGlyAlaAlaArgspMetAla 307
|||||
851 TGGCTGCTGTGGACACAGTGCCTGTGGACCCGCGGAGATATGGCC 900
308 ArgAlaGlyValSerIleProGluIleMetGlnAlaGlyLysTrpThrAs 324
|||||
901 CGCGCTGGAGTTTCATACCGGAGATCATGCAAGCTGGTGGCGACCAA 950
324 nValAsnIleValMetAsnTyrIleArgAsnLeuAspSerGluThrGlyA 341
|||||
951 TGTAAATATTGTGCATGAACATATATCCGTAACTGGATGTAAGAACAGGG 1000
341 laMetValArgLeuLeuGluAspGlyAsp 350
|||||
1001 CAATGTGCGCTGTGTGAAAGATGGCGAT 1029

seq_name: /cgn2_6/prodata/2/lna/6B.COMB.seq:US-09-193-503B-8

seq documentation block:
; Sequence 8, Application US/09193503B
; Patent NO. 6262341
; GENERAL INFORMATION:
; - APPLICANT: Baszczynski, Christopher L.
; - APPLICANT: Lyznik, Leszek A.
; - APPLICANT: Gordon-Kamm, William J.
; - APPLICANT: Guan, Xueni
; - APPLICANT: Rao, Guru
; - APPLICANT: Tagliani, Laura A.
; TITLE OF INVENTION: A No. 6262341el Method For The Integration Of Foreign DNA Into
; FILE REFERENCE: 5718-66 (amended listing)
; CURRENT APPLICATION NUMBER: US/09/193,503B
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/099,435
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: 60/056,627
; PRIOR FILING DATE: 1997-11-18
; PRIOR APPLICATION NUMBER: 60/065,613
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: encoding a FLPM:Cre polypeptide, FLP from
; OTHER INFORMATION: Saccharomyces (maize preferred codons), and Cre
; OTHER INFORMATION: from Bacteriophage P1
; NAME/KEY: CDS
; LOCATION: (1)..(2346)
; US-09-193-503B-8

alignment_scores:
      Quality: 1754.00      Length: 343
      Ratio: 5.114          Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 99.417

alignment_block:
US-09-662-128A-2 x US-09-193-503B-8 ..
Align seg 1/1 to: US-09-193-503B-8 from: 1 to: 2346

      8 ValSerAsnLeuLeuThrValHisGlnAsnLeuProAlaLeuProValAs 24
      ::::|||||
      1315 ATGGCAATTACTGACCGCTACACCAAAATTGGCTGCATTACCGGTGCA 1364
      24 palAthrSerAspGluValAlaArgLysAsnLeuMetAspMetPheArgAspA 41

```

```

|||||
1365 TGCACAGAGTGAATGAGCTTCGCAAGACCTGATGACATGTTTCAGCGATC 1414
41 rgGlnAlaPheSerGluHisThrTrpLysMetLeuLeuSerValCysArg 57
|||||
1415 GCCAGCGGTTTCTGTGAGCTACCTGGAAATCTTCTGTCCGTTGCCGG 1464
58 SerTrpAlaAlaThrPcysLysLeuAsnAsnArgLysTrpPheProAlaG1 74
|||||
1465 TCGTGGCGGCGATGTCGAAGTTGATTAACCGGAATGGTTCCGCGAGA 1514
74 uproGluAspValArgAspTyrIleuLeuTyrIleuGlnAlaArgGlyLeuA 91
|||||
1515 ACCTGAAGATGTTCCGGATTATCTTATATCTTCACGGCGCGGCTTGG 1564
91 laValLysThrIleGlnGlnHisIleuGlyGlnLeuAsnMetLeuHisArg 107
|||||
1565 CAGTAAAACTATCCACGACCATTTGGCCAGCTAACAATGCTTCATCGT 1614
108 ArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuValMe 124
|||||
1615 CGGTCCGGGCTGCCACGACCAAGTACAGCAATGCTGTTCACGTGATAT 1664
124 tArgArgIleArgLysGlnAsnValAspAlaGlyGluArgAlaLysGlnA 141
|||||
1665 GCGCGGATCCGAAAGAAAGACGTTGATGCCGCTGAACGTGCAAAACAG 1714
141 laLeuAlaPheGluArgThrAspPheAspGlnValArgSerLeuMetGlu 157
|||||
1715 CTCTAGCGTTGAAACGACGATGTTCCACAGGTCGTTCCATCAGTAA 1764
158 AsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyIleAl 174
|||||
1765 AATACCGATCGCTGCCAGATATACGTAATCTGCATTTCTGGGGATTGC 1814
174 atyAsnThrLeuLeuArgIleAlaGluIleAlaArgIleArgValLysA 191
|||||
1815 TTATTAACACCGCTGATACGTAATGCGCAGATCAGGGTTAAAG 1864
191 spIleSerArgThrAspGlyArgMetLeuIleHisIleGlyArgThr 207
|||||
1865 ATATCTCACGTACTGACGCTGGGAGAAATGTAATCCATTTGGCGAGAC 1914
208 LysThrLeuValSerThrAlaGlyValGluLysAlaLeuSerLeuGlyA 224
|||||
1915 AAACCGCTGTACGACCGCAGGTGTAGAGAAAGCACTTACCTCGGGGT 1964
224 LThrLysLeuValGluArgTrpIleSerValSerClyValAlaAspAsp 241
|||||
1965 AACTAAACTGTGTGAGCGATGATTTCCGCTCTGTGTACCTGATGATC 2014
241 roAsnAsnTyrIleuPheCysArgValArgLysAsnGlyValAlaAlaPro 257
|||||
2015 CGAATACACTACCTGTTTTCGCGGTCCAGAAAATGGTGTGGCGCGCA 2064
258 SerAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheGluAl 274
|||||
2065 TCTGCCACACGACGACTATCACTCGCCCTGGAAGGATTTTGAAGC 2114
274 atnHisArgLeuIleTyrGlyAlaLysAspAspSerGlyGlnArgTyrL 291
|||||
2115 AACTCATCGATTGATTACGGCGCTAAGATGACTCTGCTGCAGAGATACC 2164
2165 TGGCTGCTGTGGACACAGTGCCTGTGGACCCGCGGAGATATGGCC 2214
291 euA1arPserGlyHisSerAlaArgValGlyAlaAlaArgspMetAla 307
|||||
308 ArgAlaGlyValSerIleProGluIleMetGlnAlaGlyLysTrpThrAs 324
|||||
2215 CGCGCTGGAGTTTCATATACCGGAGATCATGCAAGCTGGTGGCTGCACCA 2264
324 nValAsnIleValMetAsnTyrIleArgAsnLeuAspSerGluThrGlyA 341
|||||

```

```
2265 TGTAAATTTGTCTAGTAATATCCGTAACTTGATAGTGAACAGGG 2314
341 laMeYlaJArGLeuLeuGluAspGlyAsp 350
|||||
2315 CAATGTCGCGCTGCTGGAGATGGCCAT 2343

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-735-609-11

seq_documentation_block:
; Sequence 11, Application US/08735609
; Patent No. 595360
;
GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Harligan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735,609
; FILING DATE:
; CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-735-609-11

alignment_scores:
Quality: 671.00 Length: 127
Ratio: 5.283 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-662-128a-2 x US-08-735-609-11 ..

Align seg 1/1 to: US-08-735-609-11 from: 1 to: 699

1 MetProLysLysLysArgLysValSerAsnLeuLeuThrValHisGlnAs 17
|||||
318 ATCCCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 367
|||||
17 nLeuProAlaLeuProValAlaThrSerAspGluValArgLysAsnL 34
|||||
368 TTTCCTGCGCTTACCGCTGATGCAACGAGTGTGAGTTCCGCAAGAAC 417
|||||
34 euMeLAspMeLpHeArgAspArgGlnAlaPheSerGlnHisThrTyrLys 50
|||||

2265 TGAATGACATGTTTCAGGGATCCGAGCGCTTTTCTGAGCATACCTGAAA 467
51 MetLeuLeuSerValCysArgSerTyrPAlaIaIaTrrCysLysLeuAsnAs 67
|||||
468 ATGCTTCTGTCCGTTTCCCGGTCGTCGGCGGCGCATGTGTCAAGTGAATPA 517
|||||
67 nArgLysTrrPheProAlaGluProGluAspValArgAspTyrLeuLeuT 84
|||||
518 CCGGAAATGTTTCCCGCAGAACTGAAGATGTTGGCATTTATCTTCTAT 567
|||||
84 YLeuGlnAlaArgGlyLeuValAlaValLysTrrLieglnHisLeuGly 100
|||||
568 ATCTTCAGCGCGCGGCTGCGCAGTAAATACTATCCACCAACATTTGGGC 617
|||||
101 GlnLeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspse 117
|||||
618 CAGTAAACATGCTTCATCGTCGGTCCGGGTCCGACAGACCAAGTGCACAG 667
|||||
117 rAsnAlaValSerLeuValMetArgArgIle 127
|||||
668 CAATGCTGTTTCACCTGTTATGCGCGGATC 698

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-735-609-11

seq_documentation_block:
; Sequence 11, Application US/08735609
; Patent No. 5994132
;
GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Harligan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735,609
; FILING DATE: 23-Oct-1996
; CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-735-609-11

alignment_scores:
Quality: 671.00 Length: 127
Ratio: 5.283 Gaps: 0
```


101 GlnLeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspse 117
|||||
618 CAGCTAAACATGCTTTCATGCTCGCTGGGCTCCACAGACAGTGTGAC 657

US-09-245-497-1

1 MetProLysLysLysArgLysValSerAsnLeuLeuThrValHis

34 eumetaspmetpheargasparglnalapheserglnisthtrprlys 50

```

|||||
418 TGATGACATGTCAGGATCGCAGGCGTTTCTGAGCATACCTGAA 467
51 MetLeuLeuSerValCysArgSerTrpAlaIleTrpCysLeuAsn 67
468 ATGCTTCTGTCGTTTCCGCTGCTGGCGCATGGTGCAGATTGAAATA 517
67 nArgLysTrpPheProAlaGluProGluAspValArgAspTyrLeuLeu 84
518 CCGGAATAGCTTCCGAGACCTGAAGATGTCGCGATTATCTCTAT 567
84 TyrLeuGlnAlaArgGlyLeuAlaValAlaYsthrIleGlnGlnIstleuGly 100
568 ATCTTCAGCGCGCGGCTCTGGCAGTAAACATACGACACATTTG6GC 617
618 CAGCTAAACATGCTTCATCGTCGCTCCGCGCTGCCAGACCAAGTGACAG 667
117 AsnAlaValSerLeuValMetArgArgIle 127
668 CAATGCTGTTCACTGATTATCGCGCGATC 698

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-390-51
seq_documentation_block:
: Sequence 51, Application US/09105390
: Patent No. 6288303
: GENERAL INFORMATION:
: APPLICANT: Rodriguez, Raymond
: TITLE OF INVENTION: Rice Beta-Glucanase Enzymes
: NUMBER OF INVENTION: and Genes
: NUMBER OF SEQUENCES: 66
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dehlinger & Associates
: STREET: 350 Cambridge Ave., Suite 250
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/105,390
: FILING DATE: Filed herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/050,675
: FILING DATE: 25-JUN-97
: ATTORNEY/AGENT INFORMATION:
: NAME: Pelthory, Joanne R.
: REGISTRATION NUMBER: P42,995
: REFERENCE/DOCKET NUMBER: 2000-0455.30
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-0880
: TELEFAX: 650-324-0960
: TELEX:
: INFORMATION FOR SEQ ID NO: 51:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1008 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 1...1008
: OTHER INFORMATION:
: US-09-105-390-51

```

```

alignment_scores:
  Quality: 103.50      Length: 332
  Ratio: 0.719        Gaps: 16
  Percent Similarity: 43.373      Percent Identity: 21.687

alignment_block:
US-09-662-128A-2 x US-09-105-390-51 ..

Align seg 1/1 to: US-09-105-390-51 from: 1 to: 1008

48 ThrTrpLysMetLeuLeuSerValCysArgSerTrpAlaIleTrpCysLys 64
   :::::|||||
51 TCTTGAGATTTCATTCATCCCTGAGGCGTGCATCCGTTGGTGTGT 100
   ||| :::::|||||
64 sLeu...AsnAsnArgLysTrpPheProAlaGluProGluAspValArg 80
   ||| :::::|||||
101 GCTACGGCATATCGCAGACAGATCTCCCGCGAA..... 134
80 sPtyrLeuLeuTyrLeuGlnAlaArgGlyLeuAlaValAlaYsthrIleGln 96
   ||| :::::|||||
135 .....GAGCAGCTGCTGCAGCTTCACCAATCCAA 164
97 GlnHisLeuGlyGlnLeuAsnMetLeuHisArgArgSerGlyLeuProAr 113
   ||| :::::|||||
165 TGGCATTCACAGACATGCGCATCTTACCTGCCAGCTGAGGCCATGACG 214
113 gProSerAspSerAsn.....AlaValSerLeuValMetArgArgIle 127
   ||| :::::|||||
215 CCTTCGCGGGCAGACAGCATCGGCGCTCATCGCGCGTCCCAACGACATC 264
119 .....AlaValSerLeuValMetArgArgIle 127
265 CTCATCGACCTCGCCGCCACCCGCGCTCGCGCTGCGGTGCGACGC 314
128 ArgLysGluAsnValAspAlaGly.....GluArgAlaLys 139
   :::::|||||
315 GACGCTCAGCGCGTTCGTCGCGCGGTGACATGACATGACATGCGACATGC 364
139 sGln..AlaLeuAlaPheGluArgThrAspPheAspGlnValArgSerLeu 155
   ||| :::::|||||
365 GCAACGAGATCTCGCGCGAGCCGACG.....CAGAACATCTCCCGCGCT 408
156 MetGluAsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuG 172
   ||| :::::|||||
409 ATGCAGAAC..... 417
172 YlleAlaTyrAsnThrLeuLeuArgIleAlaGluIleAlaArgIleArg 189
   ||| :::::|||||
418 .....ATCAAGCGCGCGCTGCGCGCGCGAGCATCAGCGCGCTCAAGG 460
189 allysAspIleSerArgThrAspArgGlyArgMetLeuIleHisIleGly 205
   :::::|||||
461 CCTCCACGCGCGGTGAGCTAGAC.....GTC 486
206 ArgThrLysThrLeuValSerThrAlaGlyValGluValAsn...LeuSe 221
   ||| :::::|||||
487 GTCACACACAGCTTCCCGCGCGCGCGCTGCTCGCGCGCGCTCATCAT 536
221 rLeuGlyValThrLysLeuValGluArgTrpLysSerValSerGlyValA 238
   :::::|||||
537 GACGCGCGCTGCGCCAGCTC.....CTGGATGACACCGCGCGCGC 574
238 lAspAspProAsnAsnTyr...LeuPheCysArgVal...ArgLysAsn 252
   ||| :::::|||||
575 CGCTGCTCGCCACATCATCCCTATCTTGGCTACATGCGCAAAAG 624
253 GlyValAlaAlaProSerAlaThrSerGlnLeuSerThrArgAlaLeuG 269
   :::::|||||
625 GACATCAGCTCAACTAGCCAGCTTCAGCGCGCGCGCGCGCTCCCGCA 674
269 u.....GlyIlePheGlnAlaThrHisArgL 278
   :::::|||||
675 CCCCACACCGACCTGCTAGCGCAACCTGTTGACGCGCATGTGCTGACT 724

```

```

278 euilerlycylalalysaspserglylnartryleualatrpser 294
      ::::::::::: ||| ::::|
725 CCGCTACGGCGCGTGGACAAAGCCGCGC..... 753
295 G1YHISerIalArYValG1YlaAlaArGaSPMeTAlaArGa1YVa 311
754 .....GCGGCGGGGT 764
311 lserIleProGluIleMeTGlAlaG1YlYTrP..... 322
      ||||| ::::: ::|||
765 CACATC.....GTGCTCCGAGAGCGGCGCGCGCGCGGGG 808
323 .....ThraSvAlaSnIleValMeTAsnTYrIleArGaSnleu 335
809 ACTCGGCGACGATCGACATCGCGCGACCTACGTGCAGACACTG 852
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:us-09-251-645-11

```

seq_documentation_block:

```

: Sequence 11 Application US/09251645
: Patent No. 6281413
: GENERAL INFORMATION:
: APPLICANT: Kramer, Vance C.
: APPLICANT: Morgan, Michael K.
: APPLICANT: Anderson, Arne R.
: APPLICANT: Hart, Hope
: APPLICANT: Warren, Gregory W.
: APPLICANT: Dunn, Martha
: TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
: FILE REFERENCE: CGC1963/A
: CURRENT APPLICATION NUMBER: US/09/251, 645
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 11
: LENGTH: 37948
: TYPE: DNA
: ORGANISM: Photorhabdus luminescens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (15171)..(18035)
: OTHER INFORMATION: orf5
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (23768)..(31336)
: OTHER INFORMATION: hpn2
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (31393)..(35838)
: OTHER INFORMATION: orf2
US-09-251-645-11

```

alignment_scores:

```

Quality: 96.50 Length: 491
Ratio: 0.478 Gaps: 21
Percent Similarity: 41.141 Percent Identity: 19.756

```

alignment_block:

```
US-09-662-128a-2 x US-09-251-645-11 ..
```

```
Align seg 1/1 to: US-09-251-645-11 from: 1 to: 37948
```

```

1 MetProLysLysLysArgLysVal.....SerAsnLeuLeuTh 13
  :: ||| ||| ::::|
4579 CTAGAGCAAAACCAAAAGAAATTTGCAGAGCCCAAGAAATCTCGAAAA 4628
13 rValHISGlnAsnLeuProAlaLeuProValAspAlaThrSer..... 27
  ::::| ||| ||| ::::|
4629 ATTCCGGAACACGACCACTCCCAAGAGACAGAAAAAAGCCGACA 4678

```

```

27 ..... 27
4679 GAAGACTGATATAAGCTGAAGACCCAGAAAAAAGCCGAGAAAAACCTA 4728
28 AspGluValAlArgLysAsnLeu..... 34
  ||||| ||||| ||||| |||||
4729 GATGAGGTCAGGAAAAATCTGCCAAAAGCCATATCTCTTATATGCGCCG 4778
35 .....MetAspMetPheArGaSPArgG1nAlaPheSerG1nHISTrP. 49
  ::::: ::::: ||||| |||
4779 CGCTTCGCAATTAATCACTGAAACATGCGGCAATTCCTACTCATCTGG 4828
50 .....LysMetLeuLeuSerValCysArg 57
  ||||| ::::
4829 CAGATATCTGGAACCCGCGATAGAAAAATGACAGCGGCAAAATTTATGC 4878
58 SerTrPAlaAlaTrPcysLysLeuAsnArGlyStrPheProAlaG1 74
  ::||| ::::: |||
4879 AACGTGTCGGAAAAATCCGGTTACGGCTCAACCTGAATTCGCAAAAACA 4928
74 uProGluAspValArGaSPTYrLeuLeuTYrLeuG1nAlaArgG1YLeuA 91
  : ||| ||| ::::|
4929 ACCTGAACCTGTGTCCAGTATTCG..... 4953
91 laValLysThrIleGlnHISLeuG1nLeuAsnMetLeuHISArg 107
  ||||| ::::: |||||
4954 .....CAGGCGCTGGCACAACTAGCGGTGATTTACCGT 4986
108 ArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeu.... 122
  ::||| ||||| ::::: |||||
4987 TCTACCGGCGCT.....ACGAAACACCTTAAGTCTGTTGT 5024
123 .....ValMetArgGlyLeuArgLysGluAsnValA 133
  ||||| ::::: |||||
5025 GACAAAGCCGCAACACTTGTATTAAACCAAGACCCGAAACAACCTG 5074
133 sPaLaGlyGluArgAlaLysGlnAlaLeuAlaPheGluArgTrpPhe 149
  ::::: ::::: ||| |||
5075 AAACACACCCAGACATGACGTATCAACACTAATGTCTACTAACGCGTTT 5124
150 ...AspGlnValArgSerLeuMetGluAsnSerAspArgCys..... 162
  ||| ||||| ||||| |||||
5125 ACTGACTGGGTAACTCACTAGTGAAACGCGCTCTGTGTACTAACGA 5174
163 .....GlnAspIleArgAsnLeuAlaPheLeuGlyIleAlaTYrAsnT 177
  ::::: ||||| ::||| |||
5175 ATTTAAAAAAGAACATTAAACAGCAGAACTATTGGCTAAGCTTAGTAATC 5224
177 hTrLeuLeuArgIleAlaGluIleAlaArgIleArgValys...AspIle 192
  ::::: ||| ||||| ::||| |||||
5225 TTGATAAAAATCTACGTGAGCAGACCAAGATTCAGGCNAAACTGATTGG 5274
193 SerArg.....ThraSPGlyArgMetLeuIleHISIleGlyAr 206
  ||||| ::||| ||||| ::|||
5275 TCCAACTGGCCATCTATCGACAACCTATTGTCAGTGTATTAATCTCAG 5324
206 g.....ThrLysTrPLeuValSerT 213
  | ||| |||
5325 TCAATTGAACATCTCTCCACAAGGCATTTCACACAGTCAAGTATTGCA 5374
213 hTrAla.....GlyValGluLysAla 219
  ||||| :::::
5375 CCGCAAGACCTCCGCTAATACCAATGCGAAAGCCGCTGCGATA 5424
220 LeuSerLeuGlyVal.....ThrLysLeuValGluArgTrpI1 232
  ||||| ::||| ::|||
5425 TTAAACCGCGGGGTGACACCCAAAAGACTTAACGCGCTTACATGGGTTCT 5474
232 eSerValSerGlyValAlaAspAspProAsnAsnTYrLeuPheCysArgV 249
  ::::: ||| ||| ::|||
5475 GGATGAGTCTCCAGTGTGCTGCTTAAGCAACATACTATATTATTCTGATA 5524

```

```

249 alarglys..... 251
5525 ACCAAGAGTCGAGAGCAAGAAACATACATTAATAAACCGTGA 5574
252 .....AsnGlyValAlaAlaProSer 259
5575 GATCATATACATACATATTGATGATGAACCAAGTTCCGCCGACATTA 5624
259 aThSerGlnLeuSer.....Thra 266
5625 AACACACAGATCGCTGACATCGCTAGTATCCAACTATATATTAACC 5674
266 rGAlaLeu.....GluGly.....IlePheGluAlaThrHisArg 277
5675 GCGCGTTAAATAATATGAGAGATACGTGACAAAGTGCACCAAGCCGC 5724
278 LeuIleTyrglyAlaIleLysAspSerGlyGlnArgTyrlleuAlaTrpSe 294
5725 TCATCTTCACCAACTGGGATTAATACAAATTAACGCTACAGCACTGGGC 5774
294 rGly.....HisSer 298
5775 CGGTATGGCTAACTCCTTACTATCCAGAAATTACATCGATCCGACGC 5824
298 IaArgValGlyAlaAlaArgAspMet.....Ala 307
5825 TACGTATGGCGACAGCAAAATGATGATACGTTGCTGCATTCATCAGC 5874
308 ATGAlaGlyValSerIleProGluIleMetGlnAla..... 319
5875 CAAGACCAATTAAATATGATACCTAGAAAGATGCTTTAAATCTTACT 5924
320 .....GlyGlyTrpThrAsnValAsnIleValMetAsnTyrlleu 333
5925 AACATCATTCGAACAGGTGGCTAATCTGAAATCCTCAGCCCTACCATG 5974
333 rGAsnLeuAspSerGluThrGly 340
5975 ACACACTTAATATGATCAAGGA 5997

seq_name: /cgn2_6/ptodata/2/ina/6a_COMB.seq:US-09-096-942-1
seq_documentation_block:
Sequence 1, Application US/09096942
Patent No. 6027925
GENERAL INFORMATION:
APPLICANT: Pollock, Thomas J
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
APPLICANT: Armentrout, Richard W
TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT FILING DATE: 1998-06-12
CURRENT APPLICATION NUMBER: US/09/096,942
EARLIER APPLICATION NUMBER: 60/049,428
EARLIER FILING DATE: 1997-06-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 16075
TYPE: DNA
ORGANISM: Xanthomonas campestris
PUBLICATION INFORMATION:
US-09-096-942-1

```

```

alignment_scores:
Quality: 95.50 Length: 317
Ratio: 0.628 Gaps: 18
Percent Similarity: 47.950 Percent Identity: 22.713

```

```

alignment_block:
US-09-662-128a-2 x US-09-096-942-1 ..
Align seq 1/1 to: US-09-096-942-1 from: 1 to: 16075

```

```

54 SerValCysArgSerTrpAlaAlaTrpCysLysLeuAsnAsnArgLysTr 70
4010 TCGGCGCTGTGGTTCGTGGCGGCTGTGT.....CGCTG 4044
70 PheProAlaGluProGluAspValaArgAspTyrlleuLeuTyrlleuGln 87
4045 GTGGCCCGACGACACTTG...CTGGGTGGCTTCATATCACTTCGCA 4090
87 IaArgGlyLeuAlaVallys..... 93
4091 CGCAGGGCGGTGATGTCCACGCTGTGTGTACTGGGCGCTCATCCG 4140
94 .....ThrIleGlnGlnHisLeuGlyGln.....LeuS 103
4141 GTGATGAGATCACTCATTTACTGTAGCCGTAAATCCCTGGTGGCGATGA 4190
103 nMetLeu.....HisArg 108
4191 CATGGTTGGCTATTCCGACCGCGTACATCTGGCGGTGGCGGACAGC 4240
108 rGserGlyLeuProArgProSerAspSerAsnAlaValSerLeuValMet 124
4241 GCCAGGGCTCCCGGCTGGGTGATCCGATGAGTATCGATCGATCGAC 4290
125 ArgArgIleArgLysGluAsnVal.....AspAlaGlyGluArg 137
4291 AAGAACCAACAGGTGGAGAGGTGTGATCTCGTCCGCTGGCGACGC 4340
137 gAlaLysGlnAlaLeuAlaPheGluArgTyrlleuAspPheAspGlnValArgS 154
4341 C.....GACCACATCAAGC 4354
154 rLeuMetGluAsnSerAspArg.....CysGlnAspIleArg 166
4355 AACTGCTCGACGCGCTGATCGTACCCGATCAACGTGAAGCTGTGCC 4404
167 AsnLeuAlaPheLeuGlyIleAlaTyrlleuAsnThrLeuLeuArgIleAlaG 183
4405 GACCTGTTCGACTTGGCCTGTGAACAGCTCCGACGATGGCGAG 4454
183 uIleAlaArgIleArgValLysAspIleSerArgThrAspGlyGlyArgm 200
4455 CGTGGCGGTGATCAACTGCGT.....CAGGGTGGC... 4485
200 rLeuIleHisIleGlyArgTyrlleuValSerThrAlaGlyVal 216
4486 .....GTGATCGTGAACACTTCTGTG.....GTC 4512
217 GluLysAlaLeuSer.....LeuGlyValThrLysLeuValGluArg 230
4513 GCCAAGCGCTGACGAGCAAGATCCTGGCGGTGATGGCTGATGGGCTT 4562
230 gTrp.....IleSerValSerGlyValAlaAspAspPro 242
4563 GTGGCCGCTGATCTGGCCATTCGGTA...GGCGTAAGATGACCTGGC 4609
242 snAsnTyrlleuPheCysArgValaArgLysAsnGlyValAlaAlaProSe 258
4610 CCGGCGCGGTTCCTTCGTCACGCGCGACGCGCTGGGTGGCGGAG 4659
258 rAlaThrSerGlnLeuSerThrArgAlaLeuGluGlyIlePheGluAlaTr 275
4660 TTCTACATGTTCAAGTTCGCTCGATGC..... 4687
275 hrHisArgLeuIleTyrglyAlaIleLysAspSerGlyGlnArgTyrlleu 291
4688 .....GGGTGATGACGATCATGACACGACGATTC... 4717

```

137	gAllysslnlAlaLeuAlaPheClnlgtThrAspPheAspGlnValArgS	154
4341GACCACATCAACG	4352
154	erLeuMetGlnAsnSerAspArg.....CysGlnAspLeuArg	166
4355	AGCTGCTGCACCGCTGGATCGCTACCCATCAACTGAAGCTGGTGGCC	4400
167	AsnLeuAlaPheLeuGlnGlyIleAlaTyrAsnThrLeuLeuArgIleAlaG	183
4405	GACCTGTTGCACCTCGGCGCTGTGAACAACAGTCTGCCGACACATATCGCAG	4455
183	uIleAlaArgIleArgValAllysAspIleSerArgThrAspGlnGlyArgm	200
4455	CGTCCGCGTATCAACTGGCT.....CAGGGTGCC....	4485
200	etLeuIleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyVal	216
4486GGGATCGTGACAACTACTGGTG.....GTC	4512
217	GlnLysAlaLeuSer.....LeuGlyAlaThrLysLeuValGlnArg	230
4513	GCCAAAGCGCTGCAGACAAAGATCCTCGGCGGATTTGGCTATGGGCTT	4562
230	gTTP.....IleSerValSerGlyValAlaAspAspProA	242
4563	GTCGCGCGCTGATCTGGCATATGGGTA...GGCTGAAGATGAAGCTGC	4609
242	snaSnIlePheCysArgValAlaGlyAsnGlyValAlaAlaProSe	258
4610	CGGCGCGGCTTCTTCGCTCAAGCGCGCCACAGCGCTGGGCGCGCAG	4659
258	ralaThrSerGlnLeuSerThrArgAlaLeuGlnGlyIlePheGlnAlaT	275
4660	TTCTCAATGTTCAAGTCCGCTCGATGC.....	4687
275	hnlAsArgLeuIleTyrGlnAlaLysAspAspSerGlnAlaTyrLeu	291
4688GGGTCAATGACGATATGGCACACAGTTC...G	4717
292	AlaTPSerGlnHisSerAlaTyrValAlaAlaAlaTyrAspMetAlaT	308
4718ACGACGGCAGCAAGACAGACAGCGGATTAACCGC	4752
308	gAlaGlyValSerIleProGlnIleMetGlnAlaGlyGlyTTPThrAsn	324
4753	TTCCGACAGTTTC.....TGGCGCGACAGCGCTGGACGAGC	4789

125 TGGACATGTCAGGATCGCCAGCGCTTTCTTGACCATACCTCGAANAATG 174

52 LeuLeuSerValCysArgSerTrpAlaIleArgCysLysLeuAsnAsnAr 68
 |||||
 175 CTCTCTCCGTTGCCGGTCGTGGGGCATGTGCAAGTTGAATTAACCG 224

68 gLySTPpHePrIoAlaGluPProGluAspValArgAspTyrLeuLeuTyrL 85
 |||||
 225 GAATGATTTCCCGCAGAACCTGAAGATGTGCGATTAATCTTCTATATATC 274

85 euGlnAlaArgLysLeuAlaValLysThrIleGlnGlnHisLeuGlyGln 101
 |||||
 275 TTCAGGGCGCGCTGCGCATTAATAATATCCAGACATTTGGCGCCAG 324

102 LeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAspSerAs 118
 |||||
 325 CTAAACATGCTTCATCGTCGTCGGGCTCCGACGACCAAGTACAGCAG 374

118 nAlaValSerLeuValMetArgArgIleArgLysGluAsnValAspAlaG 135
 |||||
 375 TGTGTTTCACTGCTTATGCGCGCATCCGAAAAAGAAAGCTTGATGCCG 424

135 IyGluArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheAspGln 151
 |||||
 425 GTGAACGTGCAAAACAGGCTCTACCGCTTCGACGACACTGATTTGAC 474

152 ValArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArgAsnLe 168
 |||||
 475 GTTCGTTCTACTCATGAGAAATAGCGATCGCTCCAGAGATATACGTATCT 524

168 uAlaPheLeuGlyIleAlaTyrAsnThrLeuLeuArgIleAlaGlnIleA 185
 |||||
 525 GGCATTTCTGGGAGATTGCTTATACACCGCTGTACGATAGCGCAAAATG 574

185 IaArgIleArgValLysAspIleSerArgThrAspGlyArgMetLeu 201
 |||||
 575 CCAGATCAGGGGTTAAAGATATCTCAGCTACTGACGGTGGGAGATGTGA 624

202 IleHisIleGlyArgThrLysThrLeuValSerThrAlaGlyValGlyL 218
 |||||
 625 ATCCATATATGGCAGAACGAAACGCTGTAGCACCGAGGTTGTAGAGAA 674

218 sAlaLeuSerLeu 222
 |||||
 675 GGCACCTTACCGCTG 687

seq_name: gb_est1:AW211323

seq_documentation_block:
 LOCUS AW211323 431 bp mRNA linear EST 03-DEC-1999
 u079f01.y1 NCI CGAP Mam3 Mus musculus cDNA clone IMAGE:2648761 5'
 DEFINITION similar to SW:RECR_BP1 P06956 RECOMBINASE CRE. [1] ;, mRNA
 sequence.

ACCESSION AW211323
 VERSION AW211323.1 GI:6517271
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 431)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Other_BESTs: u079f01.x1
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LIML at:
 www-bio.llnl.gov/dbfp/image/image.html

MGI:1029213
 Seq primer: -40RP from Gibco
 High quality sequence stop: 279.
 Location/Qualifiers

FEATURES

source

1..431
 /organism="Mus musculus"
 /strain="129 - C57/B6 - FVB"
 /db_xref="taxon:10090"
 /clone="IMAGE:2648761"
 /clone_id="NCI-CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /dev_stage="10 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lohar Hennighausen/Chu-Xia Deng, NIH
 Reference for transgenic model: Xu et al., Nature Genetics
 22, 37-43 (1999)."

BASE COUNT 118 a 101 c 111 g 101 t
 ORIGIN

alignment_scores:
 Quality: 652.00 Length: 143
 Ratio: 4.759 Gaps: 0
 Percent Similarity: 95.804 Percent Identity: 90.909

alignment_block:
 us-09-662-128a-2 x AW211323 ..

Align seg 1/1 to: AW211323 from: 1 to: 431

67 AsnArgLysTrpPheProAlaGluProGluAspValArgAspTyrLeuLe 83
 |||||
 3 AACCGAATGATGTTCCCGCAGAACCTGAAAGATTTGGCATTTATCTTCT 52

83 uTyrLeuGlnAlaArgGlyLeuAlaValLysThrIleGlnGlnHisLeuG 100
 |||||
 53 ATATCTTACGGCGCGCGCTGTGCGAGTAAACATATCCACACATTTGG 102

100 IyGlnLeuAsnMetLeuHisArgArgSerGlyLeuProArgProSerAsp 116
 |||||
 103 GCCAGCTAAACATGCTCATCTCGCTCGGCTGCCAGCAGCAAGTGAC 152

117 SerAsnAlaValSerLeuValMetArgArgIleArgLysGluAsnValAs 133
 |||||
 153 AGCAATGCTGTTCACTGTTATGCGGCGGATCCGAAAGAAACGTTGA 202

133 PalagIyGluArgAlaLysGlnAlaLeuAlaPheGluArgThrAspPheA 150
 |||||
 203 TGCCGGTGAACGTGCAAAACAGGCTGTAGCGTTGCAACGCACTGATTCG 252

150 spGlnValArgSerLeuMetGluAsnSerAspArgCysGlnAspIleArg 166
 |||||
 253 ACCAGTGTGTTTCATCATGGAATAATAGCAGATGCTGCAGAGATATACG 302

167 AsnLeuAlaPheLeuGlyIleAlaTyrAsnThrLeuLeuArgIleAlaG 183
 |||||
 303 AATCTGCAATTTCTGGGATGCTTATTAACACCGGATCACTAATAGCCCG 352

183 uIleAlaArgIleArgValLysAspIleSerArgThrAspGlyArgM 200
 |||||
 353 AATGCGCAGATCAGGTTAAAGATATCTCAGCGTCTCAGCGTGGAGAGA 402

200 eTleuIleHisIleGlyArgThrLysThr 209
 |||||
 403 TGTAAATCCATATTTGGGAGAGCAAGCAAC 431

seq_name: gb_est2:Br650191

```

seq_documentation_block:
  LOCUS      BI650191          563 bp      mRNA      linear      EST 12-SEP-2001
  DEFINITION  603296283F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5336853 5',
  ACCESSION  BI650191
  VERSION    BI650191.1 GI:15564427
  KEYWORDS   EST.
  SOURCE      house mouse.
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE  1 (bases 1 to 563)
  AUTHORS    NIH-MGC http://mgi.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
  COMMENT    Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-r@mail.nih.gov
              Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
              cDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://limgc.llnl.gov
              Plate: LLM11855 row: f column: 22
              High quality sequence stop: 563.
  FEATURES
    source
      1..563
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone_image="5336853"
        /clone_lib="NIH_CGAP_Mam3"
        /tissue_type="tumor, gross tissue"
        /lab_host="DH10B"
        /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
        Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
        Average insert 2 kb. Library constructed by Life
        Technologies, catalog #12017-018. Investigators providing
        samples: Lohar Hennighausen/Chu-Xia Deng, NIH Reference
        (1999). Note: this is a NCI_CGAP Library."
  BASE COUNT  143 a      133 c      149 g      138 t
  ORIGIN
alignment_scores:
  Quality: 493.00      Length: 97
  Ratio: 5.082      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
  US-09-662-128a-2 x BI650191 ..
  Align seg 1/1 to: BI650191 from: 1 to: 563
254 VALAAlaAProSeRaLaThSeRInLeuSeRThArgAlaLeuGlucI 270
|||||
1 GTTGGCGGCGCATCTGCCACCAAGCTATCACTCCGCCCTGGAGG 50
270 YLleheGluAlaThRHisArgLeuIleTyrglyAlaLysAspAspSerg 287
|||||
51 GATTTTGAAGCAACTCATCATGATTGATTCGCGCTAAGATGACCTGT 100
287 YGlnArGTYrLeuAlaTrpSerGlyHisSerAlaArgValGlyAlaAla 303
|||||
101 GTCAGAGTACTGCTGCTGCTGACACACAGTCCCGTGGAGCCGCG 150
304 ArgSpMetAlaArgAlaGlyValSerIleProGluIleMetGlnAlaG1 320
|||||
151 CGAGATATGCGCCGCGCTGAGATTCAATACCGAGATCATGCAAGCTGG 200
320 YGlyTrpThrAsnValAsnIleValMetAsnTYrIleArgAsnLeuAspS 337

```

```

seq_name: qb_est1:AW212478
seq_documentation_block:
  LOCUS      AW212478          340 bp      mRNA      linear      EST 03-DEC-1999
  DEFINITION  uc89c09.x1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:2649712 3',
  ACCESSION  AW212478
  VERSION    AW212478.1 GI:6518565
  KEYWORDS   EST.
  SOURCE      house mouse.
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE  1 (bases 1 to 340)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
  JOURNAL    Unpublished (1997)
  COMMENT    Other ESTs: uc89c09.y1
              Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-r@mail.nih.gov
              Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
              cDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bdrp/image/image.html
  FEATURES
    source
      1..340
        /organism="Mus musculus"
        /strain="129 - C57/B6 - FVB/N"
        /db_xref="taxon:10090"
        /clone_image="2649712"
        /clone_lib="NCI_CGAP_Mam3"
        /tissue_type="tumor, gross tissue"
        /dev_stage="10 months"
        /lab_host="DH10B"
        /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
        Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
        Library constructed by Life Technologies. Investigators
        providing samples: Lohar Hennighausen/Chu-Xia Deng, NIH
        Reference for transgenic model: Xu et al., Nature Genetics
        22, 37-43 (1999)."
  BASE COUNT  75 a      100 c      79 g      86 t
  ORIGIN
alignment_scores:
  Quality: 458.50      Length: 103
  Ratio: 4.776      Gaps: 1
  Percent Similarity: 93.204      Percent Identity: 89.320
alignment_block:
  US-09-662-128a-2 x AW212478/rev ..
  Align seg 1/1 to reverse of: AW212478 from: 1 to: 340
248 ArgValArgLysAsnGlyValAlaAlaProSeRaLaThSeRInLeuSe 264
|||||
340 CGGGTCAGAAAATAATGCTGTTCGCCGCCATCTGCCACAGGCGCTACNA 291

```

```

264 rThrArgAlaLeuGluglyIlePheGluAlaThrHisArgLeuIleTyrG 281
290 CTCGGG...CTGAGAGGATTTTGAAGCACTCATCGATTGATTACG 244
281 lYAlaIyAspAspSerGlyGlnArgTyrLeuAlaTrpSerGlyHisSer 297
243 GCGCTAAGCATGATCTGTGTAGAGATACCTGGCTGTGTGAGACACAGT 194
298 AlaArgValGlyAlaAlaArgAspMetAlaArgAlaGlyValSerIlePr 314
193 GCCCGTGTGGAGCGCGGAGATATGCGCCGCTGAGATTTCATACACC 144
314 GGUlMetGlnAlaGlyIleTyrTrpHisValAsnIleValMetAsnT 331
143 GGAGATCATGCAGCTGTGCTGACCAATGTAATATTGTCATGAACT 94
331 yTlIeArgAsnLeuAspSerGluThrGlyAlaMetValArgLeuLeuGlu 347
93 ATATCCGTAACCTGGATGTAACAGGGGCAATGTGCTGCTGTGGA 44
348 AspGlyAsp 350
43 GATGGCGAT 35
seq_name: gb_est1:AW211972
seq_documentation_block: 338 bp mRNA linear EST 03-DEC-1999
LOCUS AW211972
DEFINITION u09f01.x1 NCI CGAP Mam3 Mus musculus CDNA clone IMAGE:2648761 3'
similar to SW:RCCR_BPPI P06956 RECOMBINASE CRE. [1] ;, mRNA
sequence.
ACCESSION AW211972
VERSION AW211972.1 GI:6517961
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 338)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other ESTs: u079f01.y1
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.femail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnhi.gov/btrp/image/image.html
FEATURES
Source
MGI:1029213
Seq primer: -40UP from Gibco
High quality sequence stop: 151.
Location/Qualifiers
1..338
/organism="Mus musculus"
/strain="129 - C57/B6 - FVB/N"
/db_xref="taxon:10090"
/clone_id="IMAGE:2648761"
/clone_lib="NCI CGAP Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics

```

```

BASE COUNT      22, 37-43 (1999).
ORIGIN          78 a 93 c 82 g 85 t

```

```

alignment_scores:
  Quality: 397.50      Length: 103
  Ratio: 4.321         Gaps: 2
  Percent Similarity: 89.320  Percent Identity: 87.379

```

```

alignment_block:
US-09-662-128A-2 x AW211972/rev ..

```

```

Align seg 1/1 to reverse of: AW211972 from: 1 to: 338

```

```

249 ValArgLysAsnGlyValAlaAlaPro.SerAlaThrSerGlnLeuSerT 265
|||||
338 GTCCAAACAAATGTGTGTGTCGCGCATACCTACCTG...CTGCATCA 292
265 hrArgAlaLeuGluglyIlePheGluAlaThrHisArgLeuIleTyrGly 281
|||||
291 CTCGGCGCCCTGGAAGGAT.TTCTAAGCCACTCATCGATTGATTACGCG 243
282 AlaIyAspAspSerGlyGlnArgTyrLeuAlaTrpSerGlyHisSerAl 298
|||||
242 GCTAAGGATGACTTGTGTACAGATACCTGCTGTGTGACACAGATGC 193
298 aArgValGlyAlaAlaArgAspMetAlaArgAlaGlyValSerIlePrG 315
|||||
192 CCGTGTGGAGCGCGGATTTCATACATGCGCGCGGATTTCATACCGG 143
315 lUleMetGlnAlaGlyIleTyrTrpHisValAsnIleValMetAsnTyr 331
|||||
142 AGATCATGCAGCTGTGCTGACCAATGTAATATTGTCATTAACATAT 93
332 lIeArgAsnLeuAspSerGluThrGlyAlaMetValArgLeuLeuGluAs 348
|||||
92 ATCCGTAACCTGGATGTAACAGGGGCAATGTGCTGCTGTGGAAGA 43
348 pGlyAsp 350
42 TGCGGAT 36
seq_name: gb_est1:AW210918
seq_documentation_block: 343 bp mRNA linear EST 03-DEC-1999
LOCUS AW210918
DEFINITION u089c09.y1 NCI CGAP Mam3 Mus musculus CDNA clone IMAGE:2649712 5'
similar to SW:RCCR_BPPI P06956 RECOMBINASE CRE. [1] ;, mRNA
sequence.
ACCESSION AW210918
VERSION AW210918.1 GI:6516858
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 343)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other ESTs: u089c09.x1
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.femail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnhi.gov/btrp/image/image.html

```

MGI:1030164
 Seq primer: -40RP from Gibco
 High quality sequence stop: 222..
 Location/Qualifiers
 source

1.343
 /organism="Mus musculus"
 /strain="129 - C57/B6 - FVBN"
 /db_xref="taxon:10090"
 /clone_1lb="IMAGE:2649712"
 /clone_1lb="MCI CGAP Mam3"
 /tissue_type="tumor, gross tissue"
 /dev_stage="10 months"
 /lab_host="D110B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
 Reference for transgenic model: Xu et al., Nature Genetics
 22, 37-43 (1999)."
 BASE COUNT 95 a 77 c 85 g 84 t 2 others
 ORIGIN

alignment_scores:
 Quality: 393.50 Length: 115
 Ratio: 3.896 Gaps: 2
 Percent Similarity: 87.826 Percent Identity: 80.000

alignment_block:
 US-09-662-128a-2 x AM210918 ..

Align seg 1/1 to: AM210918 from: 1 to: 343

```

107 ArgArgSerGlyLeuProArgProSerAspSerAsnAlaValSerLeuVa 123
|||||
2 CGTGGGTCGGGGCTGCCACGACGACGACGACGACGATCTCTTCTCCTGCT 51
|||||
123 lMeArArgIleArgIleGlySerValAspAlaGlyGluArgAlaLysG 140
|||||
52 TATCCGGCGGATCCGAAAGAAAGCTTGATGCCGGTGAAACGTGCAAAAC 101
|||||
140 lAlaLeuAlaPheGluArgThrAspPheArgIleValArgSerLeuMet 156
|||||
102 AGGCTCAGCGTTGCAACGACATCTTCCAC.CAGGTCCTTCACTCATG 150
|||||
157 GluAsnSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyI 173
|||||
151 GAAATATAGCATCGCTCCAGATATACGTAATCTGCGCATTTGGGGGAT 200
|||||
173 eAlaTyrAsnThrLeuLeuArgIleAlaLulIleAlaArgIleArg... 188
|||||
201 TTGCTTATTAACCCCTGTTACGTATACCCGAAATTCACAGATCCANGG 250
|||||
189 ValLysAspIleSerArgThrAspGlyGlyArgMetLeuIleHisIleG 205
|||||
251 TTTAAAGATCTCTTACGCTTGACGTGGAGAAAAAATGTAATCCATATG 300
|||||
205 YArgThrLysThrLeuValSerThrAlaGlyValGluLysAla 219
|||||
301 CAGAAACAAAACCCCTGTTACCCGCCAGCTGTAGAAAGGCC 343

```

seq_name: gb_gse:FR0013439

seq_documentation_block:

LOCUS FR0013439 288 bp DNA linear GSS 18-SEP-1997
 DEFINITION F.rubripes GSS sequence, clone 122014bg11, genomic survey sequence.
 ACCESSION AL004689
 VERSION AL004689.1 GI:2450259
 KEYWORDS GSS: genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.
 1 (bases 1 to 288)
 Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umranta, Y.,
 Williams, G., and Brenner, S.
 Direct Submission
 Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
 Centre Hinxton, Cambridge, CB10 1SB. Email: biolhelp@hmp.mrc.ac.uk
 Vector: pBluescript II KS
 V-type: phagemid
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES
 source
 1..288
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_1lb="cosmid 122014"
 /clone_1lb="122014bg11"

BASE COUNT 72 a 70 c 70 g 73 t 3 others
 ORIGIN

alignment_scores:
 Quality: 261.50 Length: 78
 Ratio: 4.218 Gaps: 6
 Percent Similarity: 79.487 Percent Identity: 78.205

alignment_block:
 US-09-662-128a-2 x FR0013439 ..

Align seg 1/1 to: FR0013439 from: 1 to: 288

```

1 MetProLysLysLysArgLysValSerAsnLeuLeuThrValHisGlnAs 17
|||||
54 ATGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 103
|||||
17 neuProAlaLeuProValAspAlaThrSerAspGluValAlaGlyAsnL 34
|||||
104 TTTCCTGCTTACCGTCCGTCGATGACAGAGATGATGAGTCCCAAGAAC 153
|||||
34 euMetAspMet.PheArgAspArgGln...AlaPheSerGlnHisThrTr 49
|||||
154 TGATGACATGTTTCAGAGGATTCGCCAGGCGTTTTCGACACATACCTG 203
|||||
49 pLys...MetLeuLeuSerValCysArg...SerTrpAlaAlaTrp.Cys 63
|||||
204 GAAATATGCTTCCTGTTCCGTTTCCCGGTCGCTGGCGGCGATGTTGC 253
|||||
64 LysLeuAsnAsnArgLysTrpPhePro 72
|||||
254 AANTTTGAATTAACCCGAAATGCTTCC 281

```

seq_name: gb_est2:B1556329

seq_documentation_block:

LOCUS B1556329 836 bp mRNA linear EST 05-SEP-2001
 DEFINITION 60323744F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5290619 5',
 mRNA sequence.
 ACCESSION B1556329
 VERSION B1556329.1 GI:15443643
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 836)
 NIH-MGC <http://mgc.ncl.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 EMAIL: cga@bbs.femail.nih.gov

COMMENT

Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu

Plate: 97 row: B column: 12
Seq primer: T7
Class: BAC ends

FEATURES

High quality sequence stop: 869.

Source

Location/Qualifiers

1. 869
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACs 6; BAC Clones in E-Coli DH10B"
BASE COUNT 285 a 149 c 199 g 236 t
ORIGIN

alignment_scores:

Quality: 99.50 Length: 234
Ratio: 0.858 Gaps: 11
Percent Similarity: 49.573 Percent Identity: 24.786

alignment_block:

US-09-662-128a-2 x AZ204748 ..

Align seg 1/1 to: AZ204748 from: 1 to: 869

```

82 LeuLeuTYrLeuGlnAlaArgLyLeuAlaValLysThrIleGlnIle 98
:::|||||
75 GTGCTGTATACGCCATTAAGTGGCTACTACTTA..... 110
98 sLeuGlyIleLeuAsnMetLeuHisArgArgSerGlyLeuProArgPro 115
:::|||||
111 ..... 131
115 erAspSerAsnAlaValSerLeuValMetArgAlaGlyLeuGlnAsn 131
:::|||||
132 .....ACATGCTTTTGTAAACAATT...TGTGGCAAT 161
132 ValAspAlaGlyLeuArgAlaValGlnAlaLeuAlaPheGluArgThrAs 148
:::|||||
162 CTGCCAAACGGCAATGCACGCAAAAAGTAAACAGAAAGATCCAGTTAC 211
148 pPheAspGlnValArgSerLeuMetGluAsnSerAspArgCysGlnAsp 165
:::|||||
212 TTCAGATCAAGTTAAAGCATCTAAGAACATGTTCAATACAGAGAC. 260
165 LeuArgAsnLeuAlaPheLeuGlyIleAlaTyrAsnThrLeuLeuArgIle 181
:::|||||
261 ..GCAATTTTGTCTTCACGGG.....TTTCTAGATAT 293
182 AlaGluIleAlaArgIleArgValLysAspIleSerArgThrAspGlyG 198
:::|||||
294 GAGAACACAGATCTTGAAGGTGAATGATGTCAATTTATGAT...GA 340
198 YArgMetLeuIleHisIleGlyArgThrLysThrLeuValSerThrAla 215
:::|||||
341 TTATATTTCTTAATATTCCTAAGAGTAAGACAGATCAGTATAGACAAG 390
215 LyValGluLysAlaLeuSerLeuGlyValThr.....Lys 226
:::|||||
391 GTATATAGTATGATCATTAAGTCCGGGTGTGACAGAACTTGTCGGTCAAA 440
227 LeuValGluArgTrpIleSerValSerGlyValAlaAspAspProAsn 243
:::|||||
441 ATGTAAAGAAATATCAATGATGCGCTCGGATTTCTTTAGATTCACATCA 490

```

seq_name: gb-gss:BH093306

seq_documentation_block:

LOCUS BH093306 787 bp DNA linear GSS 18-JUL-2001
DEFINITION RPI-24-270M12.TV RPI-24 Mus musculus genomic clone RPI-24-270M12
, DNA sequence.
ACCESSION BH093306
VERSION BH093306.1 GI:14913211
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 787)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Atinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregorjls,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPI-24
Unpublished (1999)
Other-GSSs: RPI-24-270M12.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
plate: 270 row: M column: 12
Seq primer: T7
Class: BAC ends.

TITLE
JOURNAL
COMMENTFEATURES
Source

Location/Qualifiers

1. 787
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPI-24-270M12"
/clone_lib="RPI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pPARBAC1. Site 1: BamHI. Site 2: BamHI;
RPI-24 mouse BAC library produced by Pieter de Jong. The
library was cloned in the pPARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 254 a 161 c 178 g 194 t
ORIGIN

alignment_scores:

Quality: 96.00 Length: 97
Ratio: 1.778 Gaps: 3
Percent Similarity: 55.670 Percent Identity: 28.866

alignment_block:

US-09-662-128A-2 x BH093306 ..

Align seg 1/1 to: BH093306 from: 1 to: 787

```
3 LysTylsLysArgLysValSerAsnLeuThrValHisGln.AsnLeuP 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
417 AAAAAAGTGAATTTAGCTGGGCGAGTGGTGGTGCACCACTTTAAATCC 466
19 roAlaLeuProValAspAlaThrSerAsp.....GluValArgLys 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
467 CCGCACTCAAGAGGTAGAGGAGGACGACATCTTTGAGTTCAAGCCACTT 516
33 AsnLeuMetAspMetLpheArgAspArgGlnAlaPheSerGluHisThr.. 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
517 GGCTACAAAGCAAGTTCAGAGACAGCCAGTGTACACAGAGAAACCTT 566
49 .....TrpLysMetLeuLeuSerValLysArgSerT 59
567 AATGATATACAAATAGAGGCTGAGCTACGCCCTGCAATATGATGACAGAT 616
59 rPAlaAlaIrrPCysLysLeuAsnAsnArgLysTrpPheProAlaGluPro 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
617 GGGCAGCTGGTGGTTCATGACAGTCCGCCCAAAACACACAGCAGCAATCCT 666
76 GluAspValArgAspTyrLeuLeuTyrLeuGlnAlaArg 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
667 GCACCCCAATGAGCACCTGTTGTGGCCCGCTGGGAGCA 705
```

seq_name: qb_gss:AG083458

seq_documentation_block:

LOCUS AG083458 1287 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-080013.R, genomic survey sequence.
ACCESSION AG083458
VERSION AG083458.1 GI:16635260
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone:lib.PTB Chimpanzee Male
BAC library clone:PTB-080013.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE

AUTHORS 1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE

JOURNAL BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 1287)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.

REFERENCE

AUTHORS Direct Submission
Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpesgsc.riken.go.jp, Fax:81-45-503-9170)
Tel:81-45-503-9111, Fax:81-45-503-9170
Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

COMMENT

PRIMERS

Sequencing: M13rev

LIBRARY

Vector : PK5145
R.site 1 : SacI
R.site 2 : SacI
Location/Qualifiers
1..1287

/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-080013.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 222 a 569 c 379 g 99 t 18 others
ORIGIN

alignment_scores:

Quality: 96.00 Length: 226
Ratio: 0.865 Gaps: 10
Percent Similarity: 49.115 Percent Identity: 25.664

alignment_block:

US-09-662-128A-2 x AG083458/rev ..

Align seg 1/1 to reverse of: AG083458 from: 1 to: 1287

```
142 LeuAlaPheGluArgThrAspPheAspGlnValArgSerLeuMetGlns 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
873 CTGTGCTGTGTGGCGGTGACATCGGGCGCGCGCGCG.....TCGCG 830
158 nSerAspArgCysGlnAspIleArgAsnLeuAlaPheLeuGlyIleAlaT 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
829 GCGGAGCGGGTGGCGACGTGCGGGCTTCGCGGAGCTGGGGCTCTCTG. 781
175 rAsnThrLeuLeuArgIleAlaGlnIleAlaArgIleArgVal..... 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
780 .....GCGCTTCGCGCGTGGCGCGCGTCCGCTCGGTGCGGTGTC 736
190 .....LysAspIleSerArgThrAspG 197
735 GTGGCAGTCTGCGCTGCGGGGAGTACCGGGCGGTGCGGGGGGGGG 686
197 yGluArgMetLeuIleHisIleGlyArgThrLysThrLeuValSerThra 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
685 CAGCAGG.....GTGGCCGCCACCGCACCTGTGTCGGCGGAGG 642
214 IaGlyValGluLysAlaLeuSerLeuGlyValThrLysLeuValGluArg 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
641 CCGGGGTGGCGCGTGTCTGTACTC..... 616
231 TrpIleSerValSerGlyValAlaAspAspProAsnAsnTyrLeuPheCy 247
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
615 ...ATGAGTCTGCGCGCGCTGTTCGGACTATCGGTAGTGGTCTGCGG 569
247 sArgValArgLysAsn.....GlyValA 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
568 GCGCCGCGCGGGCGCGCTCGGGGGGGGGCGAGGGGCTCGGCTCGG 519
255 IaAlaProSerAlaThrSerGlnLeuSerThrArgAlaLeuGluIle 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
518 GCGGCGGGGGGGCGCGCGGGGGGGGGCGGCGCACAGGGCGCCGATA 469
272 PheGluAlaThrHisArgLeuIleTyrGlyAlaLysAspAspSer..... 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
468 GCGGCGCGGGGGGTGCGGTAAATGTGCGGTGCGTGAAGCCGCGGG 419
287 .GlyGlnArgTyrLeuAlaTrpSerGly..... 295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 GGGTCTGCGGTGGTGGCGCGGTGCGGCGAGTGGCGGGGGCGGGTGG 369
296 .....HisSerAlaArgValGlyAla 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
368 CACCTACAGCTTCATTCGCGGGGCGCTCACATCAGTACAGTGGGGGCG 319
303 AlaArgAspMetAlaArgAlaGlyValSerIlePro.....GluI 316
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318 GCGCGG.....GTGGCAGCGGGGGTGGCGGGGCTGTTGGTGGCGGCGT 272
316 eMetGlnAlaGlyLysTyrThrAsnVal 325
```

```

      seq_name: gb_est2:BF797551
      271 GGTGGCGCGGGGGGCTGCTGATGTG 244
      :::: |||||:::
      seq_documentation_block:
      LOCUS BF797551 778 bp mRNA linear EST 12-JAN-2001
      DEFINITION 602257335F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4340720 5',
      mRNA sequence.
      ACCESSION BF797551
      VERSION BF797551.1 GI:12102605
      KEYWORDS EST
      SOURCE human.
      ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      NIH-MGC http://mgc.nci.nih.gov/.
      National Institutes of Health, Mammalian Gene Collection (MGC)
      Unpublished (1999)
      CONTACT: Robert Strausberg, Ph.D.
      Email: cga@bbs-remail.nih.gov
      Tissue Procurement: Louis Staudt, M.D., Ph.D.
      cDNA Library Preparation: Life Technologies, Inc.
      CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLM953 row: e column: 09
      High quality sequence stop: 749.
      Location/Qualifiers
      1..778
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:4340720"
      /clone_lib="NIH_MGC_85"
      /tissue_type="lymphoma, cell line"
      /lab_host="DH10B (phage-resistant)"
      /note="Organ: lymph. Vector: pCMV-SPORT6, site_1: NotI;
      site_2: SalI; Cloned unidirectionally; oligo-dT primed.
      Average insert size 1.867 kb. Library enriched for
      full-length clones and constructed by Life Technologies.
      Note: this is a NIH_MGC library."
      BASE COUNT 159 a 225 c 230 g 164 t
      ORIGIN
      alignment_scores:
      Quality: 94.50 Length: 198
      Ratio: 0.936 Gaps: 8
      Percent Similarity: 51.010 Percent Identity: 27.273
      alignment_block:
      US-09-662-128a-2 x BF797551/rev ..
      Align seg 1/1 to reverse of: BF797551 from: 1 to: 778
      160 AsparGcGlnAspIleArg.....AsnLeuAl 169
      |||||
      672 GACAGGTCGACGACGACGCTGGGCTCGCTGTCAAGACAAATCTTAA 623
      169 apheLeuGlyIleAlaTyrasnThrLeuLeuArgIleAlaGluIleAla 186
      |||
      622 AGGAGCTGCGCAAGCTCGAGACGAGGAGATGACGTCGACGAACTCAGACA 573
      186 rglleArGValIysAspIleSerArgThrAspGlyGlyArgMetLeu... 201
      |||
      572 GGTGATGAGTTCATCCCTGTCAATGTCAAGACTCTCCGAGATGTTGTGCG 523
      202 .....IleHisIleGlyArgThrIleValSerThr..... 213
      |||
      522 ATGAGCTGCTTCATCTCAGACGACGACACTAAGCCGTGTGTCTCGCCCTTCC 473

```

```

      214 .....AlaGlyValGluValAlaLeu 221
      472 CGTAGGCACTTACCAACCGGCTCAGGTCCTTCTGTTCAAGGTTCCCT 423
      221 erLeuGlyValThrLysLeuValGluArg...TrpIleSerValSerGly 236
      |||
      422 CATCATCAAAAGTCAAGATGCGGAGGACATTAATGGACTTGAATGTCGGC 373
      237 ValAlaAspAspProAsnAsnIleTyrLeuPheCysArgValArgLysAsnG1 253
      |||||
      372 GTGGCTGTGTCACTGAACACA.....CTGAGAGATCCAGAGAACTCTTC 329
      253 Y.....ValAlaAlaProSerAlaThrSerGlnLeuSerThrA 266
      |||
      328 AAGGCTAAGGCTGTCTTGGCTGGGATGTGAGAAACCCGCAAGATTC 279
      266 rglAlaLeuGluGlyIlePheGluAlaThrHisArgLeuIleTyrGlyAla 282
      |||
      278 GCTCTTGAAGGGGTTGGCTTGACCTGTGAAAGGCTGAGAATCTGCTCG 229
      283 Lys.....AspAspSerGlyAlaArgTyrLeuAlaTr 293
      |||
      228 AAGGCACTTGTGCCGAGTAGCAGATCCACGCTCCGCTCTCTGCGG 179
      293 pSer.GlyHisSerAlaArgValGlyAlaAlaArgAspMetAlaArgAla 309
      |||
      178 AAGCACTCACAACAAACCGCTGTGGGTAGGAGATCTCTGCTCTGCTCA 129
      310 GlyValSerIleProGluIleMetGlnAlaGlyGlyTrpThr 323
      |||
      128 GGAGCTCAAGTCTGTGCTACTCGCGCCAGCAGCTCTCTTGACACA 87
      seq_name: gb_est2:BG629464
      seq_documentation_block:
      LOCUS BG629464 894 bp mRNA linear EST 22-MAY-2001
      DEFINITION 602763759F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4899062 5',
      mRNA sequence.
      ACCESSION BG629464
      VERSION BG629464.1 GI:14177051
      KEYWORDS EST.
      SOURCE human.
      ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      NIH-MGC http://mgc.nci.nih.gov/.
      National Institutes of Health, Mammalian Gene Collection (MGC)
      Unpublished (1999)
      CONTACT: Robert Strausberg, Ph.D.
      Email: cga@bbs-remail.nih.gov
      Tissue Procurement: ATCC
      cDNA Library Preparation: Ling Hong/Rubin Laboratory
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLM1790 row: e column: 15
      High quality sequence stop: 866.
      Location/Qualifiers
      1..894
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:4899062"
      /clone_lib="NIH_MGC_42"
      /tissue_type="epithelioid carcinoma cell line"
      /lab_host="DH10B (phage-resistant)"
      /note="Organ: pancreas. Vector: pCMV7, site_1: XhoI;
      site_2: EcoRI; cDNA made by oligo-dT priming.
      directionally cloned into EcoRI/XhoI sites using the
      following 5' adaptor: GGCACGAG(5). Size-selected >500bp

```



```

602 GACAGGTGCCAGACACAGC.....TGGGGCTGCTGCACAG 565
176 nhrleuLeuArgIleAlaGluIleAla..... 185
564 GACAACTTTAAAG.....GAGCTGGCAAAAGTGGAGAACGGAGATGA 521
186 .....ArgIleArgValIysAspIleSerArgThrAsp 196
520 CGTGTGGAACACAGAGAGTTGATGGTTCATCCCTGTCATATGCACAC 471
197 GtGtGtArgMetLeu.....IleHisIleGtIleArgThrLysThrLe 210
470 TCTCTCAGAGTGTTCGATGAGTGTTCATCTCAGACGACACTAAGCG 421
210 vAlSerThr..... 215
420 TGTGTCTCCGCCCTCTCCCGTGGAGCACTTACACAGCGGCTCAGGTCTT 371
215 lYValGluLysAlaLeuSerLeuGtIleValThrLysLeuValGluArg... 230
370 CTCTGTTCAGAGTTCCTCATCATCAAGTCAAGATGCGAGAGGCATTA 321
231 TrpIleSerValSerGtIleValAlaAspAspProAsnSerIleuPheCy 247
320 TGGGACTGTGATGTGGCGGTGGCTGTCTACTGAACACA.....CTGAG 277
247 sArgValArgLysAsnGly.....ValAlaAlaProSerAlaT 260
276 GAGATCCAGAGAACGCTCAAGGCTGTCTTGGCTGGGATGAG 227
260 hrSerGlnLeuSerThrArgAlaLeuGtIlePheGluAlaThrHis 276
226 AGAAGACCTGCAGATTGCTCTTGAAAGGGTGGCTTGAAGCTTGA 177
277 ArgLeuIleTyrGlyAlaLys.....AspAspSerG 287
176 AGGCTGGAATCTGCTCAAGGAGCACTGTGCTCCGAAGTGCAGCTCCAC 127
287 yGlnArgTyrLeuAlaTrpSer..GlyHisSerAlaArgValGtIleAla 303
126 GCTCCGCTGCTCTCGGGAGACAGCTCACAAAACCGCTGGCTGAGA 77
304 ArgAspMetAlaArgAlaGtIleValSerIleProGluIleMetGlnAlaG 320
76 GCATCTCTCTGCTCTGAGGACGTCAAGTCTGTACTCGGCCAGCAGC 27
320 yGltTyrThr 323
26 TCTTGGACA 17
seq_name: gb_gss:A2145302
seq_documentation_block:
LOCUS A2145302 453 bp DNA linear GSS 28-AUG-2000
DEFINITION SP_0044.B1.F05.spe6 Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate-44 Col-9 Row-L, DNA sequence.
ACCESSION A2145302
VERSION A2145302.1 GI:8297205
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 453)
Cameron,R.A., Mahalinas,G., Rast,J.P., Martinez,P., Bondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Gray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and additional resources
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

```

```

MEDLINE 20402566
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 44 row: L column: 9
Seq primer: Sp6
Class: BAC ends
High quality sequence stop: 453.
FEATURES
source location/Qualifiers
1..453
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone_lib="Plate-44 Col-9 Row-L"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACs.6; BAC Clones in E-Coli DH10B"
BASE COUNT 138 a 102 c 114 g 97 t 2 others
ORIGIN
alignment_scores:
Quality: 92.50 Length: 164
Ratio: 1.088 Gaps: 10
Percent Similarity: 51.829 Percent Identity: 28.049
alignment_block:
US-09-662-128a-2 x A2145302 ..
Align seg 1/1 to: A2145302 from: 1 to: 453
184 lLeAlaArgIleArgValIysAspIleSerArgThrAspGtIleArgme 200
26 ATAGCAGAGGATGATCAAAATTTTGCAGTGGCCGCAAGCAAGC...CT 72
200 lLeuIleHisIleGtIleArgThrLysThrLeuValSerThrAlaGtIleG 217
73 GCTAGTACGTATTGATTTCTCAAAAC...GACCAATGACGGCCAAAG 116
217 lLysAlaLeuSerLeuGly.....ValThr 225
117 GAACCAATCATATCGGTAGCAATGGACAGATGAAGAAGTCCCTGTG 166
226 lLysLeuValGluArgTrpIleSerValSerGtIleValAlaAspProAs 242
167 AAAGCCATGAAGACATACCTCCGTGAAGACCTGATATGATGATGCCG.. 214
242 nasnTyrLeuPheCyArgValArgLysAsnGtIleValAlaAlaProSer 259
215 .....TATTATTGT.....CATCGTATGGG.....GCCCTGTGA 245
259 lAthSerGlnLeuSerThrArgAlaLeuGtIlePheGluAlaThr 275
246 CTAGGTATCAATTTAGTTCA.....CTTCTGAAGAAAC 280
276 HisArgLeuIleTyrGtIleValAlaLysAspAspSerGtIleArgTyrLeuAl 292
281 CTACAC.....CACAAAGGTACAGTGTGATCCCAATAT... 313
292 atpPserGlyHisSerAlaArgValGtIleValAlaArgAspMetAlaArg 309
314 ....GGCACACACCTCTTTCGAATGGAGCGGCAACACACAGCGGCTGA 359
309 lAgGtIleValSerIleProGluIleMetGlnAlaGtIleTyrThrAsnVal 325
360 AAGGCGTTCGCCNACAGACATATGCAATGGGAAGTGGCGGAGC... 406
326 asnIleValMetAsnTyrIleArgAsnLeuAspSerGtIleThr 339
.....

```

407 TCTGTACATGCGTTATATCCGCTAAATATCAGGTGCACT 448
